

AD-A118 214

NAVAL POSTGRADUATE SCHOOL MONTEREY CA

F/G 12/1

A GRAPHICAL TEST BED FOR ANALYZING AND REPORTING THE RESULTS OF--ETC(U)

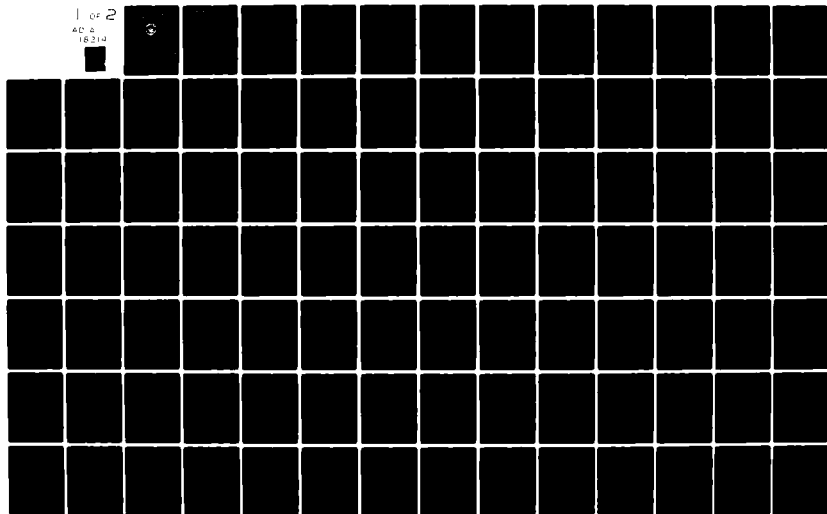
MAR 82 D 6 LINNEBUR

UNCLASSIFIED

NL

1 of 2

AD-A  
16 214



AD A118214

# NAVAL POSTGRADUATE SCHOOL

Monterey, California

(2)



## THESIS

A GRAPHICAL TEST BED  
FOR ANALYZING AND REPORTING  
THE RESULTS OF A  
SIMULATION EXPERIMENT

by

David George Linnebur

March 1982

Thesis Advisor:

P. A. W. Lewis

DTIC  
ELECTE

16 1982

A

Approved for public release; distribution unlimited

82 08 16 170

DTIC FILE COPY

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	AD A118214	
4. TITLE (and Subtitle)		5. TYPE OF REPORT & PERIOD COVERED
A Graphical Test Bed for Analyzing and Reporting the Results of a Simulation Experiment		Master's Thesis March 1982
7. AUTHOR(s)		6. PERFORMING ORG. REPORT NUMBER
David George Linnebur		
9. PERFORMING ORGANIZATION NAME AND ADDRESS		8. CONTRACT OR GRANT NUMBER(s)
Naval Postgraduate School Monterey, California 93940		
11. CONTROLLING OFFICE NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Naval Postgraduate School Monterey, California 93940		
13. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE
		March 1982
		13. NUMBER OF PAGES
		112
		14. SECURITY CLASS. (of this report)
		Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)		
Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
Simulation, graphics, box plots, regression, output analysis, bias reduction		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)		
A graphical test bed in which the results of a simulation experiment can be reported and analyzed is described. The test bed is based on the regression adjusted graphics and estimation (RAGE) methodology developed by Heidelberger and Lewis [Ref.1] for regenerative simulations. From the graphics and the associated numerics the experimenter can summarize and see		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE  
S/N 0102-014-6001

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE/When Data Entered

simultaneously relative properties, such as bias, normality and standard deviation, of several estimators of a characteristic of a population for up to eight sample sizes. The graphics is supported on a line printer to make it and the program portable.



Accession	✓
Index	
Classification	
Distribution	
Availability Codes	
Avail and/or	
Dist	Special

A

Approved for public release; distribution unlimited

A Graphical Test Bed  
For Analyzing and Reporting  
The Results of a  
Simulation Experiment

by

David George Linnebur  
Captain, United States Marine Corps  
B.A., Saint Mary of the Plains College, 1975

Submitted in partial fulfillment of the  
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL  
March 1982

Author:

David G. Linnebur

Approved by:

P.C.W. Lewis  
Thesis Advisor

J.P. [unclear]  
Second Reader

[unclear]  
Chairman, Department of Operations Research

W.M. Woods  
Dean of Information and Policy Sciences

## ABSTRACT

A graphical test bed in which the results of a simulation experiment can be reported and analyzed is described. The test bed is based on the regression adjusted graphics and estimation (RAGE) methodology developed by Heidelberger and Lewis [Ref. 1] for regenerative simulations. From the graphics and the associated numerics the experimenter can summarize and see simultaneously relative properties, such as bias, normality and standard deviation, of several estimators of a characteristic of a population for up to eight sample sizes. The graphics is supported on a line printer to make it and the program portable.

## TABLE OF CONTENTS

I.	INTRODUCTION .....	11
II.	USAGE .....	14
III.	DESCRIPTION OF RAGE ARGUMENTS .....	24
	A. DATA ARRAY---X .....	24
	B. SAMPLE SIZE---N .....	24
	C. NUMBER OF REPLICATIONS---M .....	25
	D. SUBSAMPLE SIZE ARRAY---NE .....	25
	E. NUMBER OF SUBSAMPLE SIZES (Box Plots)---L ---	26
	F. DEGREE OF REGRESSION---D .....	27
	G. REDUCED GRAPHICS---RG .....	27
	H. SCALING ESTIMATORS INDIVIDUALLY---SEI .....	28
	I. SETTING THE VERTICAL SCALE---SVS .....	28
	J. MINIMUM VALUE OF THE VERTICAL SCALE---YMIN --	29
	K. MAXIMUM VALUE OF THE VERTICAL SCALE---YMAX --	29
	L. NUMBER OF ESTIMATORS---NEST .....	29
	M. ESTIMATOR FUNCTIONS---EST1, EST2, EST3 .....	30
	N. DESCRIPTION OF ESTIMATORS---TTL1, TTL2, TTL3 -	30
IV.	EXAMPLE OF A PROGRAM USING RAGE .....	32
V.	A COMPARATIVE STUDY OF ESTIMATORS FOR THE GAMMA SHAPE PARAMETER .....	35
	A. EXAMING THE GAMMA DISTRIBUTION FOR THE SHAPE PARAMETER (K=5.0) .....	38
	1. Case 1, Example 1: Small Subsample Sizes and Small Replications (K=5.0) .....	38

2.	Case 1, Example 2: Small Subsample Sizes and Large Replications (K=5.0)	47
3.	Case 1, Example 3: Large Subsample Sizes and Large Replications (K=5.0)	61
B.	EXAMING THE GAMMA DISTRIBUTION FOR THE SHAPE PARAMETER (K=0.2)	70
1.	Case 2, Example 1: Small Subsample Sizes and Small Replications (K=0.2)	70
2.	Case 2, Example 2: Small Subsample Sizes and Large Replications (K=0.2)	77
3.	Case 2, Example 3: Large Subsample Sizes and Large Replications (K=0.2)	86
	RAGE PROGRAM LISTING	93
	LIST OF REFERENCES	110
	INITIAL DISTRIBUTION LIST	111

# LIST OF TABLES

1.	Standard deviations of estimates of the shape parameter using the moment and MLE estimators for small sample sizes (replications=2) -----	39
2.	Skewness coefficients of the estimates of the shape parameter using the moment and jackknifed moment estimators (replications=2) -----	40
3.	Skewness coefficients of the estimates of the shape parameter using the MLE and jackknifed MLE estimators (replications=2) -----	40
4.	Standard deviations of the estimates of the shape parameter using the moment and MLE estimators for varying sample sizes (replications=100) -----	48
5.	Bias coefficients for each estimator's population of estimates and their standard deviations for small and large replications (M=2 and M=100) -----	52
6.	Standard deviations of the population of estimates of the shape parameter for each estimator for subsample sizes 500 and 2000 -----	62
7.	Bias and variance comparisons for varying subsample sizes and shape parameters -----	88

## LIST OF FIGURES

1.	M replications, each of sample size N -----	15
2.	Subsectioning MxN samples into subsamples of size $n_k$ , $k=1, \dots, K$ -----	16
3.	(a, b, c, d) Full scale graphics of 4 estimators (Moment, MLE, Jackknifed Moment and Jackknifed MLE) estimating the shape parameter ( $K=5.0$ ) of the gamma distribution with small subsamples (33-500) and small replications (2) -----	43
4.	(a, b, c, d) Full scale graphics of 4 estimators (Moment, MLE, Jackknifed Moment and Jackknifed MLE) estimating the shape parameter ( $K=5.0$ ) of the gamma distribution with small subsamples (33-500) and large replications (100) -----	53
5.	(a, b, c, d) Reduced graphics of 4 estimators (Moment, MLE, Jackknifed Moment and Jackknifed MLE) estimating the shape parameter ( $K=5.0$ ) of the gamma distribution with small subsamples (33-500) and large replications (100) -----	57
6.	(a, b, c, d) Full scale graphics of 4 estimators (Moment, MLE, Jackknifed Moment and Jackknifed MLE) estimating the shape parameter ( $K=5.0$ ) of the gamma distribution with large subsamples (100-2000) and large replications (25) -----	64
7.	(a, b) Reduced graphics of 2 estimators (Moment and MLE) estimating the shape parameter ( $K=5.0$ ) of the gamma distribution with large subsamples (100-2000) and large replications (100) -----	68
8.	(a, b, c, d) Full scale graphics of 4 estimators (Moment, MLE, Jackknifed Moment and Jackknifed MLE) estimating the shape parameter ( $K=0.2$ ) of the gamma distribution with small subsamples (33-500) and small replications (2) -----	73
9.	(a, b, c, d) Full scale graphics of 4 estimators (Moment, MLE, Jackknifed Moment and Jackknifed MLE) estimating the shape parameter ( $K=5.0$ ) of the gamma distribution with small subsamples (33-500) and large replications (100) -----	80

10. (a, b) Reduced graphics of 2 estimators (Moment and MLE) estimating the shape parameter ( $K=0.2$ ) of the gamma distribution with small subsamples (33-500) and large replications (100) ----- 84
11. (a, b, c, d) Full scale graphics of 4 estimators (Moment, MLE, Jackknifed Moment and Jackknifed MLE) estimating the shape parameter ( $K=0.2$ ) of the gamma distribution with large subsamples (100-2000) and large replications (25) ----- 89

## ACKNOWLEDGEMENT

I wish to acknowledge two people that contributed to my thesis research. First is Professor P. A. W. Lewis who was my thesis advisor. His time was always available and his advice and support contributed a great deal to my knowledge and the completion of this thesis for which I am grateful.

Second I would like to thank Professor Luis C. Uribe who was the primary contributor of the graphics provided by the "RAGE" FORTRAN program which is a part of this thesis. Those changes and additions that were made were frequently with his advice and assistance and contributed significantly to the final program.

## I. INTRODUCTION

During the course of a simulation experiment the experimenter is frequently confronted with the problem of estimating some function or parameter from the generated data. He will usually have a choice of one or more estimators, parametric or nonparametric, that can be used to estimate the function or parameter. These estimators will vary in their statistical properties. Some will estimate with low variance but contain bias, others will have just the opposite characteristics and still others will contain varying degrees of both. How these different estimators compare in relation to each other as well as over varying sample sizes are of concern to the experimenter. Comparison of estimators are important in that they help the experimenter to choose the most appropriate estimator for his particular analysis. Comparisons may also be helpful in deciding whether a new estimator is of any use. The appropriate estimator to use is a function of the bias, variance and nonnormality in the estimate for available run length (sample size).

When the simulated data from which the properties of the estimators is to be examined is independent and identically distributed (iid) it can be sectioned into  $M$  independent blocks of specified sample size ( $N$ ). From each of these blocks the estimator can be applied to all subsamples of

size  $n_k$ , thereby creating samples of estimates of varying size for each estimator. As a first cut at the properties of the estimators, box plots of these samples of estimates are drawn for subsample sizes  $n_k$  (see, e.g. Figure 3a).

However, something is usually known about the properties of the estimators and these properties can be used to enhance the graph of box plots versus subsample size,  $n_k$ . Thus it is often known that bias in the estimator is a polynomial in  $1/n$ , and the variance of the estimator is a polynomial in  $\frac{1}{n}$ ,  $\frac{1}{n^{1.5}}$ ,  $\frac{1}{n^2}$ , ... Also the estimators are usually asymptotically normally distributed. Consequently the graphs are enhanced by an estimated unbiased regression line for the mean of the estimates of the parameter as well as a horizontal line for the regression-estimated expected value of the parameter (asymptote). A regression is also performed for the variance structure of the estimators and the sample characteristics (kurtosis and skewness) are given under each boxplot to assist in judging the convergence to normality.

Heidelberger and Lewis [Ref. 1] used this methodology in the context of regenerative simulation to develop a procedure and protocol to produce unbiased and normally distributed regression adjusted regenerative estimates (RAGE), with a given precision or for total available computing time. However, the graphics associated with this procedure required a Tektronic 4013 terminal using an APL graphics package making

its use limited to those organizations possessing such a terminal. The protocol also requires a relatively experienced user to properly interpret the graphical and statistical output and to apply the sequential steps of the protocol.

The purpose of this thesis is to apply the methodology developed by Heidelberger and Lewis [Ref. 1] to comparing the bias and variance of estimators of a parameter from simulated output. This objective also includes supporting the graphics associated with the methodology on a line printer to make its use more accessible.

Details of the graphics and methodology are given in Chapter II, entitled USAGE. Details of the arguments of the RAGE program are given in Chapter III and a sample program using the RAGE graphics is given in Chapter IV. All the user has to supply is the input data and FORTRAN functions to compute the estimators for any given subset of data.

A comparative study of estimators of the gamma shape parameter using the graphical test bed RAGE is given in Chapter V. This comparative study illustrates the different properties of estimators that can be seen with regression adjusted graphics as sample size increases.

## II. USAGE

The regression adjusted graphics and estimating (RAGE) program is designed to evaluate distributional properties of one or more estimators as a function of sample size from a stationary series of raw data. These estimators are programmed as functions by the user and declared external in the main program. The graphical and numerical results that accompany each estimator are varied and will depend to some extent on the options desired by the user. The purpose of the following discussion is to explain the graphical and numerical output as well as how they were obtained and some of the options available to the user.

The general procedure begins with a given iid data base or one that has been simulated and an unknown parameter or population characteristic,  $\theta$ , that is to be estimated by an estimator,  $\hat{\theta}$ , from the data. The experimenter must first decide on the number of independent blocks or sections ( $M$ ) of size  $N$  the data is to be sectioned into. These  $M$  sections or blocks will henceforth be referred to as replications; the terminology will seem more appropriate as the discussion continues. The number of replications ( $M$ ) can alternatively be determined indirectly by the sample size ( $N$ ) desired in each section. In either case  $M \times N$  must not exceed the total allowable sample size. This sectioning of the data into

M replications of sample size N is represented in Figure 1 where the entire block represents the MxN samples from the total sample population.

N	$m_1$	
N	$m_2$	
.	.	$m_i \quad i=1, \dots, M$
.	.	
N	$m_M$	

Figure 1: M replications, each of sample size N

Within each replication the experimenter must decide on the subsample sizes that the estimator,  $\hat{\theta}$ , is to use in estimating the parameter. The number of different subsample sizes, K, that can be used to resection the same data can vary from one to eight and the size of each subsample can vary from one to the size of the sample, N. This repeated subsectioning of the same data is represented pictorially by Figure 2 where  $[x]$  refers to the highest integer less than or equal to x and  $r_k$ ,  $k=1, \dots, K$  refers to the number of estimates possible from a sample of size N using subsample sizes  $n_k$ ,  $k=1, \dots, K$  to estimate the parameter. Each box or rectangle in Figure 2 represents the same MxN data set with the same M replications of size N subsectioned into different sections of size  $n_k$ . Note that the  $n_k$ 's divide N evenly in Figure 2. This is not necessary but certainly desirable since all available information will be utilized.

<u>Sample size</u>	<u>Total Samples MxN</u>						<u>Section (Rep.)</u>	<u>Population of Estimates for subsample size n</u>
N	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	m <sub>1</sub>	$\left[ \frac{N}{n_1} \right] \times M = r_1 M$
N	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	m <sub>2</sub>	
.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	
N	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	m <sub>M</sub>	
N	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	m <sub>1</sub>	$\left[ \frac{N}{n_2} \right] \times M = r_2 M$
N	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	m <sub>2</sub>	
.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	
N	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	n <sub>2</sub>	m <sub>M</sub>	
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
N	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	m <sub>1</sub>	$\left[ \frac{N}{n_k} \right] \times M = r_k M$
N	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	m <sub>2</sub>	
.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	
N	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	n <sub>k</sub>	m <sub>M</sub>	

Figure 2: Subsectioning MxN samples into subsamples of size n<sub>k</sub>, k=1, ..., K

If each subsample size (n<sub>k</sub>, k=1, ..., K) produces a total of r<sub>k</sub>, k=1, ..., K estimates per sample of size N then there will be a population of estimates from the M replications that total Mx r<sub>k</sub>, k=1, ..., K for each subsample. For each

population of estimates from each subsample ( $n_k$ ,  $k=1, \dots, K$ ) box plots are plotted so that the distribution of the estimates for the varying subsample sizes can be seen graphically.

In addition to these box plots of the estimates, a regression adjusted estimate (RARE) of the parameter is computed and plotted over the boxplots. In addition to the RARE estimate, unbiased RARE estimates out to a specified number of terms determined by the user are computed and plotted on the graph (The mathematical derivation of the RARE and unbiased RARE estimates are given later in this section). The unbiased RARE estimates are computed as a function of subsample size and plotted in conjunction with the RARE estimate so that a comparison of the two values can be made graphically and the evolution to an unbiased state can be seen as the subsample size increases.

This idea can be made clearer through the use of an example and its associated graphics and numerics, Figure 4a. Here an independent and identically distributed sample of 50,000 gamma distributed random numbers (shape parameter,  $K=5.0$ ) has been generated and sectioned into  $M=100$  sections (replications) with each replication containing  $N=500$  random numbers (sample size). This information is printed at the top of the graph for future identification. The experimenter chose eight different subsample sizes at which to examine the parameter (33, 50, 71, 125, 166, 250 and 500). These subsample sizes ( $n_k$ 's) are printed on the horizontal scale

of the graph. The population of estimates of the parameter at each subsample size are then plotted as vertical box plots above these numbers. When the population of estimates determined from a subsample size is less than nine, box plots are omitted and the values of these estimates are simply plotted with the symbol "O" on the graph. There are no estimates with populations totaling less than nine in Figure 4a but the graphical output for such cases can be seen in Figure 3a where subsample sizes, 125, 166, 250 and 500 total less than nine estimates.

The horizontal scale of the graph is scaled roughly to the minimum and maximum subsample sizes with the box plot of each subsample's estimates spaced proportionally. However, when subsample size differences are relatively small so that overlap of the box plots would occur, a minimum of six spaces between centers of box plots is left to prevent the graphics from becoming confused.

The vertical scale of the graph is determined by the minimum and maximum values from the population of estimates determined from the smallest subsample size ( $n_1$ ). It is assumed here that subsequent subsample sizes ( $n_2, n_3, \dots$ ) will be large in relation to the smallest subsample size ( $n_1$ ) so that estimates produced from these will be within the range of that determined from the smallest subsample size. This assumption appears tenable if the range of estimates is compared in Figure 4a for increasing subsample sizes.

The vertical scale of Figure 4a is not scaled to the minimum and maximum value of the estimates determined from the smallest subsample size ( $n_1=33$ ). This is because the RAGE program can evaluate up to three different estimators at one time and two alternate estimators (Figures 4b & 4c) were also used to estimate the parameter. An option was entered to scale all three estimators to the same scale and it chose the minimum and maximum values from all the estimates produced by all three estimators using the smallest subsample size ( $n_1=33$ ). By comparing Figures 4a, 4b and 4c it can be seen that the maximum value was taken from the first estimator, Figure 4a, and the minimum value was taken from the third estimator, Figure 4c. Other scaling options will be discussed in detail in the following chapter.

Below the graph statistical results as they pertain to the population of estimates for each subsample size are printed. These statistics include the sample mean, standard deviation, standard deviation of the mean, skewness and kurtosis. Thus, for example, in Figure 4a at subsample size  $n_1=33$  there is a population of  $[N/n_1] \times M = r_1 \times M = [500/33] \times 100 = 150$  estimates for which a sample mean, standard deviation, standard deviation of the mean, skewness and kurtosis is computed. The subsample sizes these statistics refer to are labeled across the top.

As mentioned earlier, in addition to box plots and associated statistics of the estimates, RARE and unbiased RARE

estimates of the parameter are computed and plotted on the graph. The degree of regression,  $D$ , for the computations of the unbiased RARE estimates is determined by the user and printed at the top of the graph for future reference. The RARE estimate is the regression coefficient  $\beta_0$  and is the estimated asymptotic expected value of the parameter being estimated. The unbiased RARE estimates are determined by the regression coefficients,  $\beta_1, \beta_2, \dots, \beta_d$ , the subsample size,  $n$ , and the RARE estimate,  $\beta_0$ , using the following formula:

$$\beta_0 + \beta_1 \frac{1}{n} + \beta_2 \frac{1}{n^2} + \dots + \beta_d \frac{1}{n^d}.$$

A set of regression coefficients ( $\beta_0, \beta_1, \dots, \beta_d$ ) is determined from the mean of the population of estimates (dependent variable) at each subsample size for each replication. The  $M$  regressions that determine the  $M$  sets of coefficients ( $\beta_0, \beta_1, \dots, \beta_d$ ) are represented by the following formula:

$$\begin{array}{ccccccc} {}_m\hat{\theta}(n_1) & = & \beta_0 & + & \beta_1 \frac{1}{n_1} & + & \beta_2 \frac{1}{n_1^2} + \dots + \beta_d \frac{1}{n_1^d} \\ \vdots & & \vdots & & \vdots & & \vdots \\ {}_m\hat{\theta}(n_k) & = & \beta_0 & + & \beta_1 \frac{1}{n_k} & + & \beta_2 \frac{1}{n_k^2} + \dots + \beta_d \frac{1}{n_k^d} \end{array}$$

$$m = 1, \dots, M$$

$$\text{where } \hat{\theta}_m(n_k) = \frac{\sum_{i=1}^{r_k} \hat{\theta}_i(n_k)}{r_k} \quad \begin{array}{l} k = 1, \dots, K \\ m = 1, \dots, M \end{array}$$

The regression coefficients, to determine the RARE and unbiased RARE estimates, are the average values determined from these M regressions applied to the M replications. The mean, variance and standard deviation of these coefficients are printed below the graph below the subsample size statistics.

As mentioned earlier, the first coefficient,  $\beta_0$ , is the estimated asymptotic expected value of the parameter being estimated (RARE). Its mean from the M regressions is printed as a horizontal dashed line across the graph. As the subsample size increases the bias in the unbiased RARE estimates becomes smaller and the unbiased RARE estimates (dotted line) will asymptotically approach the RARE estimate (dashed line). In Figure 4a the estimated expected value or asymptote  $\beta_0 = 4.98$  and the bias coefficients for the three degree (D) regression are:  $\beta_1 = 16.50$ ,  $\beta_2 = -182.21$ ,  $\beta_3 = 3659.04$ .

The variance regression coefficients are determined from the variances of the population of estimates for each subsample size (dependent variable) and evaluating the single regression

$$\begin{aligned}
\text{VAR}(\hat{\theta}(n_1)) &= \alpha_1 \frac{1}{n_1} + \alpha_2 \frac{1}{n_1^{1.5}} + \dots + \alpha_{d+1} \frac{1}{n_1^{(d+2)/2}} \\
&\vdots \\
\text{VAR}(\hat{\theta}(n_k)) &= \alpha_1 \frac{1}{n_k} + \alpha_2 \frac{1}{n_k^{1.5}} + \dots + \alpha_{d+1} \frac{1}{n_k^{(d+2)/2}}
\end{aligned}$$

$$\text{where } \text{VAR}(\hat{\theta}(n_k)) = \frac{\sum_{m=1}^M \sum_{i=1}^{r_k} [\hat{\theta}(n_k) - m\hat{\theta}_i(n_k)]^2}{r_k^M - 1}$$

$k=1, \dots, K$

$$\text{and } \hat{\theta}(n_k) = \frac{\sum_{m=1}^M \sum_{i=1}^{r_k} m\hat{\theta}_i(n_k)}{r_k^M}$$

$k=1, \dots, K$

The four (D+1) variance regression coefficients for Figure 4a are;  $\alpha_1 = 58.37$ ,  $\alpha_2 = 174.72$ ,  $\alpha_3 = -1500.13$ , and  $\alpha_4 = 5446.25$ . The variance regression coefficients derived here can be used to obtain an approximate variance for a specified subsample size by substituting the subsample size for  $n_k$  in any one of the above regression equations along with the variance coefficients ( $\alpha_1, \alpha_2, \dots$ ) and completing the computation.

Below the variance regression coefficients the minimum and maximum value of the vertical scale is printed. The purpose of this information is to allow the user to scale other estimators of the same parameter to the same scale. It can be used when the number of estimators exceeds three

(the maximum number of estimators RAGE can use in a single run) or when subsequent runs for different conditions are desired to be of the same scale for graphical comparisons.

A final identifying feature of the graphics is a section at the very bottom of the graphical output which allows the user to identify the estimator that was used in estimating the parameter. In Figure 4a this estimator was the moment estimator, i.e. the reciprocal of the estimated coefficient of variation squared.

### III. DESCRIPTION OF RAGE ARGUMENTS

The RAGE program has the calling sequence: call RAGE(X,N,M,NE,L,D,RG,SEI,SVS,YMIN,YMAX,NEST,EST1,TTL1,EST2,TTL2,EST3,TTL3). The arguments are relatively easy to understand but in several cases are related and must meet certain restrictions. For this reason a detailed description of each argument is given, as well as any restrictions or limitations that must be met within or between arguments. Following this detailed discussion an example is given in Chapter IV on how to set up the main program and the estimator functions.

#### A. DATA ARRAY...X

The X argument is a single precision Real\*4 array containing the data from which the parameter is to be estimated. The only restriction is that the size of the X array must not exceed 50,000 which is the maximum sample size that RAGE can store.

#### B. SAMPLE SIZE...N

The number of data elements per section or sample size is designated by the integer N. Depending on the minimum sample size required by the estimator, N can vary from 1 to 50,000. The sample size, N, times the number of sections, M, cannot exceed 50,000 (the maximum number of data elements

stored in the array X). If M times N exceeds 50,000 an error message will be printed and execution will terminate. If M times N exceeds the number of data elements stored by the user in the array X then no error message will be printed and the program will execute giving unpredictable results or an abnormal ending will occur.

#### C. NUMBER OF REPLICATIONS...M

The number of sections or replications the array X is sectioned into is determined by the integer value assigned to M. The number assigned to M is also the number of regressions that will be run to determine the regression coefficients  $\beta_0, \beta_1, \dots, \beta_d$ . The number of regressions is limited to one hundred and therefore M is limited to a maximum value of one hundred. An error message will be printed and execution will terminate if M is greater than one hundred or less than one. If M is equal to one the variance and standard deviation of the regression coefficients for the unbiased estimate will not be printed since they will not exist.

#### D. SUBSAMPLE SIZE ARRAY...NE

The argument NE is an integer array of size eight containing the subsample sizes ( $n_k, k=1, \dots, K$ ) the estimator is to use in estimating the parameter. The values in the array NE should be ordered from smallest to largest or a warning message will be issued to indicate the vertical scale of the graph may not be large enough to include all

estimates. RAGE will use the population of estimates determined from the subsample size stored in NE(1) to determine the vertical scale of the graph. If NE(1) is not the smallest value the range of the graphics may not be large enough to include outliers determined from subsequent smaller subsample sizes.

Due to storage limitations a maximum of 12,500 estimates is permissible. Since NE(1) will produce the most estimates, the user can determine the total number of estimates using the formula  $M \times [N/NE(1)]$ . If this value exceeds 12,500 an error message will result stating this limitation has been exceeded and execution will terminate. Corrective action can be taken by increasing NE(1) or decreasing M or N.

#### E. NUMBER OF SUBSAMPLE SIZES (BOX PLOTS)...L

The user can select the number of subsample sizes to be used from the array NE(8) by assigning an integer value to the argument L. The argument, L, can take on values between one and eight. RAGE will then select the first L subsample sizes from NE(8), compute the population of estimates of the parameter with each, graph the box plots and print the associated statistics. If L is not an integer value between one and eight an error message will result and execution will terminate.

If  $L = 1$  no regressions are possible and any regression output is omitted. The maximum degree of regression allowed

by RAGE, regardless of the value assigned to the argument D (degree of regression), is L-1. If D exceeds this value RAGE will set the degree of regression to L-1.

#### F. DEGREE OF REGRESSION...D

The degree of regression, D, is an integer value less than or equal to three. The degree of regression refers to the number of bias coefficients,  $\beta_1, \beta_2, \dots, \beta_D$ , that will be used in computing the unbiased estimate out to D terms.

In addition to the restrictions on D mentioned earlier, D will also be reduced for the regression that determines the variance coefficients,  $\alpha_1, \alpha_2, \dots, \alpha_{D+1}$ , when the number of subsamples sizes with a variance is less than D+1.

#### G. REDUCED GRAPHICS...RG

The reduced graphics argument, RG, allows the user to reduce the range of the vertical scale of the graph by eliminating outliers and only printing the number of outliers that exist. The vertical scale is reduced to the upper quartile plus 1.5 times the interquartile distance value and the lower quartile minus 1.5 times the interquartile distance value of the estimates from the smallest subsample size, NE(1). The vertical scale then becomes these values and the outliers are merely counted and printed on the graph at the ends of each box plot's upper (lower) quartile + (-) 1.5 times interquartile distance position. If there are no outliers for the NE(1) estimates, either or both extremes, the

vertical scale is further reduced to the first estimate value within these limits.  $RG = 1$  will cause reduced graphics to be produced and  $RG = 0$  will cause full scale graphics.

#### H. SCALING ESTIMATORS INDIVIDUALLY...SEI

The argument SEI allows the user to scale each estimator's vertical scale for the graphics separately or to the same scale.  $SEI = 0$  causes all estimators' vertical scale to have the same range of values. The range of the vertical scale and whether outliers are printed or counted will depend on the graphics desired ( $RG = 0$  or  $1$ ). Figures 4a, 4b and 4c are the results of three estimators scaled to the same scale ( $SEI = 0$ ) with full scale graphics ( $RG = 0$ ). Figures 5a, 5b and 5c are the results of three estimators scaled to the same scale ( $SEI = 0$ ) but with reduced graphics ( $RG = 1$ ).

If  $SEI = 1$  each estimator's vertical scale will be scaled individually. The range of the vertical scale and whether outliers are printed or counted will depend on the graphics desired ( $RG = 0$  or  $1$ ).

#### I. SETTING THE VERTICAL SCALE...SVS

To set the vertical scale an integer value of 1 is assigned to the argument SVS.  $SVS = 1$  will cause the SEI argument to be ignored and the vertical scale of all the estimators' graphics will be set to those values assigned to YMIN and YMAX. When the argument SVS is in effect ( $SVS = 1$ ) the argument RG is also in effect as far as whether outliers

are printed or counted ( $RG = 0$  or  $1$ ) but the determination of the vertical scale by  $RG$  is overridden by  $SVS$  and is set to  $YMIN$  and  $YMAX$ .  $SVS = 0$  causes  $RAGE$  to ignore values assigned to  $YMIN$  and  $YMAX$  and scaling is determined by the values assigned to the arguments  $RG$  and  $SEI$ . Figure 4d is an example of  $SVS = 1$  and  $RG = 0$  with the vertical scale set to that of Figures 4a, 4b, and 4c ( $YMIN = -1.2812$ ,  $YMAX = 13.9984$ ). Figure 5d is an example of  $SVS = 1$  and  $RG = 1$  with the vertical scale set to that of Figures 5a, 5b and 5c ( $YMIN = 1.0394$ ,  $YMAX = 9.1423$ ).

J. MINIMUM VALUE OF THE VERTICAL SCALE...YMIN

The value  $YMIN$  is a real number assigned by the user as the lower limit of the vertical scale. If  $YMIN$  is assigned a value that is so large that it falls within the range of values that determine the body of any box plot then an error will occur and an abnormal ending will result.

K. MAXIMUM VALUE OF THE VERTICAL SCALE...YMAX

The value  $YMAX$  is a real number assigned by the user as the upper limit of the vertical scale. If  $YMAX$  is assigned a value that is so small that it falls within the range of values that determine the body of any box plot then an error will occur and an abnormal ending will result.

L. NUMBER OF ESTIMATORS...NEST

The argument  $NEST$  is an integer, 1, 2 or 3, which indicates the number of estimators that will be used to estimate the

parameter. An error message will be issued if NEST is not one of these values and execution will terminate.

#### M. ESTIMATOR FUNCTIONS...EST1,EST2,EST3

These arguments are the function names assigned to each of the estimators that are to be used in estimating the parameter. Each function must be written with a two argument calling sequence (i.e. call EST1(X,N)) where the first argument, X, must be the data array and the second argument, N, must be the number of data elements.

If there are less than three estimators then dummy arguments must be inserted in the calling sequence where subsequent estimators would normally appear (Suggestion...use a function of a previous estimator, i.e. EST1, as the dummy. RAGE will not provide duplicate output for these repeated functions unless NEST is increased to include them.). Each function must then be declared external in the main program. For an illustration of how the functions are written and declared see the example in Chapter IV.

#### N. DESCRIPTION OF ESTIMATORS...TTL1,TTL2,TTL3

The titles associated with each estimator can be passed as an argument through the call program (RAGE) provided there is a minimum of 120 characters between apostrophes (i.e. call RAGE(..., EST1,'---120 characters---',EST2,...)). An alternate and possibly easier method is to declare each title (TTL1, TTL2, TTL3) as a Real\*8(15) array and assign

the characters that describe each estimator as data to each title. For an example of this method see the example in Chapter IV.

The user should note that the title associated with each estimator must immediately follow the estimator argument. As in the case of estimators, dummy titles must be inserted when the number of estimators requiring titles is less than three. Finally, it should be noted that the program will not execute properly unless titles of the correct format are included.

#### IV. EXAMPLE OF A PROGRAM USING RAGE

In the following example suppose a user has a set of iid data from which he wishes to estimate the population mean by using two estimators, the sample mean and the sample median. Assume that these estimators are given the abbreviated function names MN and MED and that the associated titles for each estimator are "The sample mean as an estimator of the population mean" and "The sample median as an estimator of the population mean." If the subsample sizes at which it is desired to examine the properties of these estimators are 33, 50, 71, 125, 166, 200, 250 and 500 the FORTRAN program may look as formatted on the following page.

```

REAL*4 X(50000),YMIN,YMAX
REAL*8 T1(15),T2(15)
INTEGER NEST,D,RG,SEI,SVS,M,N,L,NE(8)
DATA NE/ 33,50,71,125,166,200,250 500/
C   INSERT 8 CHARACTERS BETWEEN EACH SET OF APOSTROPHES UNTIL
C   TITLE IS COMPLETED AND FILL REMAINDER WITH BLANKS.
DATA T1/'THE SAMP','LE MEAN ','AS AN ES','TIMATOR ',
+'OF THE P','OPULATIO','N MEAN',8*' '/
DATA T2/'THE SAMP','LE MEDIA','N AS AN ','ESTIMATO',
+'R OF THE','POPULATI','ON MEAN',8*' '/
EXTERNAL MN,MED

```

.  
.  
.

```

CALL RAGE(X,N,M,NE,L,D,RG,SEI,SVS,YMIN,YMAX,
+NEST,MN,T1,MED,T2,MN,T1)

```

.  
.  
.

```

STOP
END

```

```

      REAL FUNCTION MN(X,N)
      REAL X(N)
C     AVERAGE OF N VALUES IN X
      MN=0.0
      DO 10 I=1,N
         MN = MN +X(I)
10    CONTINUE
      MN = MN/FLOAT(N)
      RETURN
      END

```

```

      REAL FUNCTION MED(X,N)
      REAL X(N)
C     MEDIAN OF N VALUES IN X
C     "PCTILE" IS A FUNCTION THAT RETURNS THE 50 PERCENTILE
C     NUMBER FROM THE DATA ARRAY X OF SIZE N.
      MED = PCTILE(X,N,.5)
      RETURN
      END

```

In this example since only two estimators were used a dummy estimator and title had to be inserted as the arguments for the third case. Here the first estimator and its title were simply repeated.

V. A COMPARATIVE STUDY OF ESTIMATORS FOR THE GAMMA SHAPE  
PARAMETER

To demonstrate the use of the Regression Adjusted Graphics and Estimating (RAGE) package, two arrays of gamma distributed random variables with shape  $K=5.0$  and  $K=0.2$  were generated using the Naval Postgraduate School's random number generator--LLRANDOMII [Ref. 3]. Because these random numbers are independent and were drawn from the same distribution, they can be sectioned into independent and identically distributed blocks of size  $N$ . RAGE was specifically designed to evaluate estimators of parameters from just such iid populations. These two samples then should demonstrate the capabilities of the RAGE package to evaluate the bias, variance and distribution of estimators of a specified parameter from the populations.

In the following examples it is the shape parameter,  $K$ , of the gamma distribution that is being estimated. There are two primary estimators used in these examples. The first is the moment estimator (reciprocal of the estimated coefficient of variation squared),  $K = \frac{1}{(C(X))^2} = \frac{\bar{X}^2}{\text{VAR}(X)}$ . The second

estimator is the maximum likelihood estimate (MLE) [Refs. 2,4] which is determined by a bisection search for the shape parameter,  $K$ , until the following equation is satisfied:

$$\ln(k) - \text{Digamma}(K) = \ln \bar{X} - \frac{\sum_{i=1}^n \ln X_i}{n} \quad i=1, \dots, n.$$

The  $X_i$ ,  $i=1, 2, \dots, n$  are the random numbers from a specified block of size  $n$ . The other two estimators investigated are jackknifed versions of the initial estimators. The jackknife routine [Ref. 5] with each of these estimators is a four fold jackknife which has the form

$$\tilde{\theta}_* = \frac{1}{4} \sum_{j=1}^4 \tilde{\theta}_{*j}$$

where  $\tilde{\theta}_{*j} = 4(\tilde{\theta}_{\text{all}}) - 3(\tilde{\theta}_j)$ ,  $\tilde{\theta}_{\text{all}}$  is the estimate from the entire sample ( $n$ ) and  $\tilde{\theta}_j$  is the estimate with the  $j_{\text{th}}$  group missing. Each group is a set of one fourth of the data points.

Each of the four estimators is examined for varying subsample sizes and replications to determine how accurately each estimator estimates the shape parameter of the gamma distribution as the subsample size increases. The shape parameter of the gamma distribution was chosen as an example because the behavior of the many estimators which have been proposed for  $K$  is fairly well known. Roughly speaking, it is known that the moment estimator is quite efficient for  $K > 2$  but quite inefficient for  $K < 1$  even for large sample sizes. Bias and normality of these estimators is not documented. Again the jackknife is known to work well with moment estimators

even in small samples but no such results are known for the jackknifed MLE estimator except that it is asymptotically efficient.

For this investigation there are basically three examples representing different simulation scenarios for evaluating each of two cases,  $K=5.0$  and  $K=0.2$ . They are:

- (a) Small subsample sizes (33 to 500), small replications ( $M=2$ );
- (b) Small subsample sizes (33 to 500), large replications ( $M=100$ );
- (c) Large subsample sizes (100 to 2000), large replications ( $M=25$ ).

The first can be looked on as a pilot study to determine regions of interest for the two longer scale studies. However, the procedures are not fully sequential, unlike the usage of the RAGE methodology proposed by Heidelberger and Lewis [Ref. 1] for regenerative simulations.

With many of these examples another set of reduced graphics is included which reduces the vertical scale and provides graphics which are more suitable for examining the bias structure of the estimators. The reduced graphics truncates the outliers in the box plots giving only their number and not their position.

A. EXAMINING THE GAMMA DISTRIBUTION FOR THE SHAPE PARAMETER  
( $K=5.0$ )

1. Case 1, Example 1: Small Subsample Sizes and  
Small Replications ( $K=5.0$ )

Figures 3a, 3b, 3c and 3d show the graphical and statistical results from estimating the shape parameter with the four different estimators. These graphs all have the same vertical scale so that meaningful comparisons can be made of the box plots and of the bias of each estimator.

One of the characteristics of the MLE estimator is that its estimates will asymptotically have the smallest variance of all estimators. For small subsample sizes (such as the case of Figures 3a & 3b) the asymptotically smaller variance of the MLE estimator is not apparent in the box plots. The body of the box plot for the MLE estimator for  $n_1 = 33$  is smaller than the same box plot for the moment estimator. However, the MLE has more and larger outliers, indicating that the MLE is more nonnormal at this sample size than the moment estimator. If, however, the estimated standard deviation of the various subsample sizes are compared for each of the two primary estimators (Table 1) it can be seen that the MLE estimator consistently produces a smaller variance than does the moment estimator. From Figures 3a & 3b the bias in the two estimators appears comparable, starting high but becoming negligible by subsample size 250.

Table 1: Standard deviations of estimates of the shape parameter using the moment and MLE estimators for small sample sizes (replications = 2)

SAMPLE SIZE	<u>Standard Deviation</u>	
	MOMENT ESTIMATOR	MLE ESTIMATOR
33	1.517	1.306
50	.9942	.7151
71	.8829	.6051
100	.6312	.4612
125	.7725	.4823
166	.8310	.5114
250	.4660	.2772
500	.5589	.3087

Jackknifing is a technique that can be used on iid samples to reduce the bias in the estimates caused by the estimator [Ref. 5]. An estimator with a complete jackknife is asymptotically normal with the same variance as the unjackknifed estimator, but small sample properties are difficult to obtain analytically. Figures 3c & 3d are the jackknifed estimators and again because of the small number of estimates it is not apparent that the jackknife is contributing to nonnormality or increasing variance when the box plots are compared with those of Figures 3a & 3b. The increased normality, however, can be seen in the jackknifed moment estimator by comparing box plots (Figures 3a & 3c) and the coefficients of skewness (Table 2) with the nonjackknifed form. The change in normality or symmetry of the estimates from the jackknifed MLE estimator

is not discernable in a comparison of the box plots (Figures 3b & 3d) or the coefficients of skewness (Table 3).

Table 2: Skewness coefficients of the estimates of the shape parameter using the moment and jackknifed moment estimators (replications = 2)

<u>Skewness Coefficients</u>		
SAMPLE SIZE	MOMENT ESTIMATOR	JACKKNIFED FORM
33	.553	.327
50	.076	-.348
71	.339	.253
100	.499	.157
125	1.573	.724
166	1.018	.501
250	-.150	-.235
500	N/A	N/A

Table 3: Skewness coefficients of the estimates of the shape parameter using the MLE and jackknifed MLE estimators (replications = 2)

<u>Skewness Coefficients</u>		
SAMPLE SIZE	MLE ESTIMATOR	JACKKNIFED FORM
33	1.198	.050
50	.510	.616
71	.193	-.163
100	.773	.643
125	1.451	1.216
166	1.101	1.229
250	-.030	.081
500	N/A	N/A

The bias in each estimator is depicted by the space between the dotted line and dashed line of each of the graphs. When the unbiased estimate (dotted line) approaches the estimated expected value (dashed line) then the bias becomes insignificant. Each estimator can be expected to vary in the amount of bias it has in its estimates and this bias can be expected to diminish as sample size increases. Comparing Figures 3a & 3b it appears that the MLE estimator has about the same bias as the moment estimator but this is hard to determine at this point since the bias coefficients were determined from a small number of estimates. In the estimates from the jackknifed estimators one would expect the bias to be reduced because jackknifing is specifically designed to remove the  $\beta_1 (1/n)$  term in the bias of the estimates but this is not obvious from the graphical comparisons (Figures 3a & 3c and Figures 3b & 3d). For the case of the MLE estimator the coefficient of the  $1/n$  term,  $\beta_1$ , has actually increased from 32.65 to 75.33 after jackknifing. The standard deviation of these bias terms are so large (19.14 and 16.23 respectively), though, that they suggest more estimates (increased replications) are necessary to better determine the bias.

Determining how much more bias one estimator has than another can also be accomplished by comparing the bias regression coefficients  $\beta_1$ ,  $\beta_2$ , etc. The larger these coefficients are the greater is the bias of that particular

estimator. Comparing the coefficients for the moment and MLE estimators of Figures 3a & 3b show that the MLE estimator has a larger bias. For a subsample size of 500 and a three degree regression the bias for the moment estimator is

$$\frac{7.40128}{500} + \frac{183.099}{500^2} + \frac{1947.86}{500^3} = .0155$$

and the bias for the MLE estimator is

$$\frac{32.6486}{500} - \frac{1998.68}{500^2} + \frac{47308.0}{500^3} = .0577$$

These comparisons seem to indicate that the MLE estimator has more bias. This is not necessarily true, though, for reasons mentioned earlier. Increasing the number of replications will alleviate much of this uncertainty since more precision in the estimates of the bias coefficients will result.

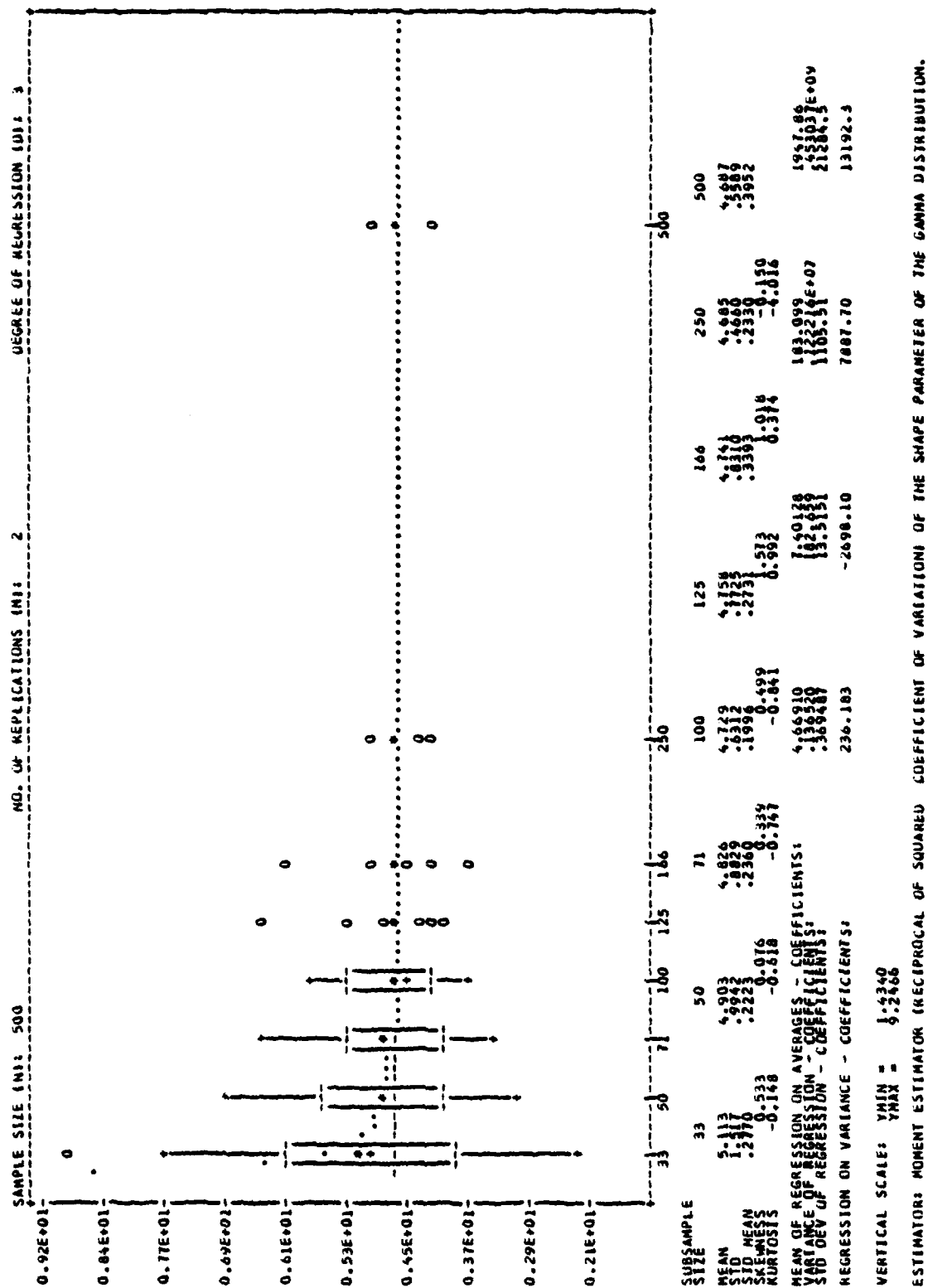
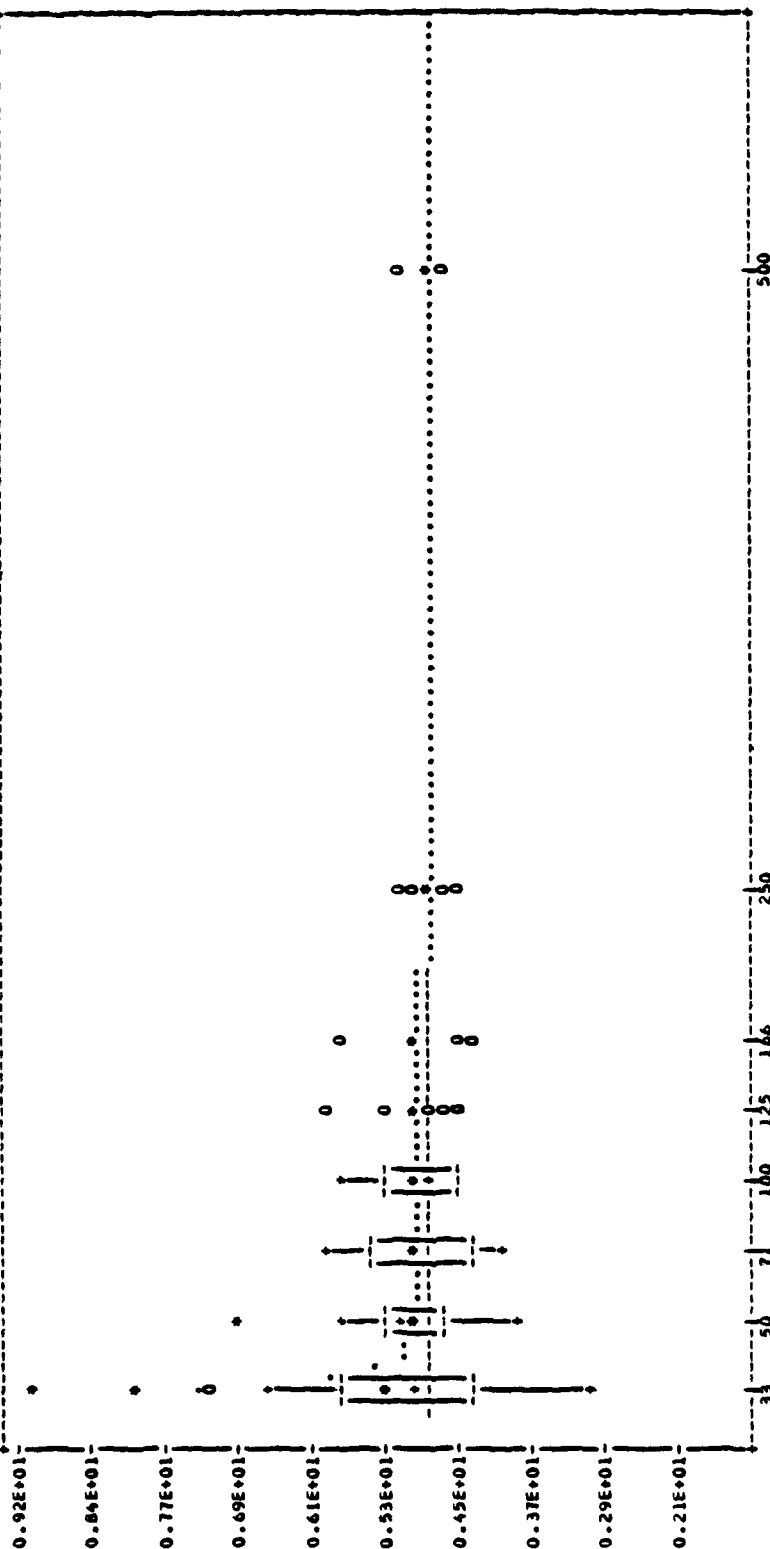


Figure 3a

SAMPLE SIZE (N): 500

NU. OF REPLICATIONS (M): 2

DEGREE OF REGRESSION (D): 3



SAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	5.289	5.049	5.021	4.990	4.994	4.963	4.997	4.990
STD MEAN	1.306	.7137	.6051	.4613	.4823	.5114	.4772	.3647
STD DEV	.2385	.1600	.1611	.1478	.1763	.2088	.1386	.2183
SKENESS	1.198	0.510	0.193	0.773	0.521	0.101	-0.040	
KURTOSIS	1.730	1.069	-1.089	-0.361	0.984	0.274	-2.475	
MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:				4.82090	32.6486		-1998.48	47308.0
VARIANCE OF REGRESSION ON AVERAGES - COEFFICIENTS:				.18225E-01	18.1349		1091.24	11309.1
STD DEV OF REGRESSION ON AVERAGES - COEFFICIENTS:				.185001				
REGRESSION ON VARIANCE - COEFFICIENTS:				-19.9361	2112.48		-25709.4	92320.9

VERTICAL SCALE: YMIN = 1.2340  
YMAX = 4.2466

ESTIMATOR: MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 3b

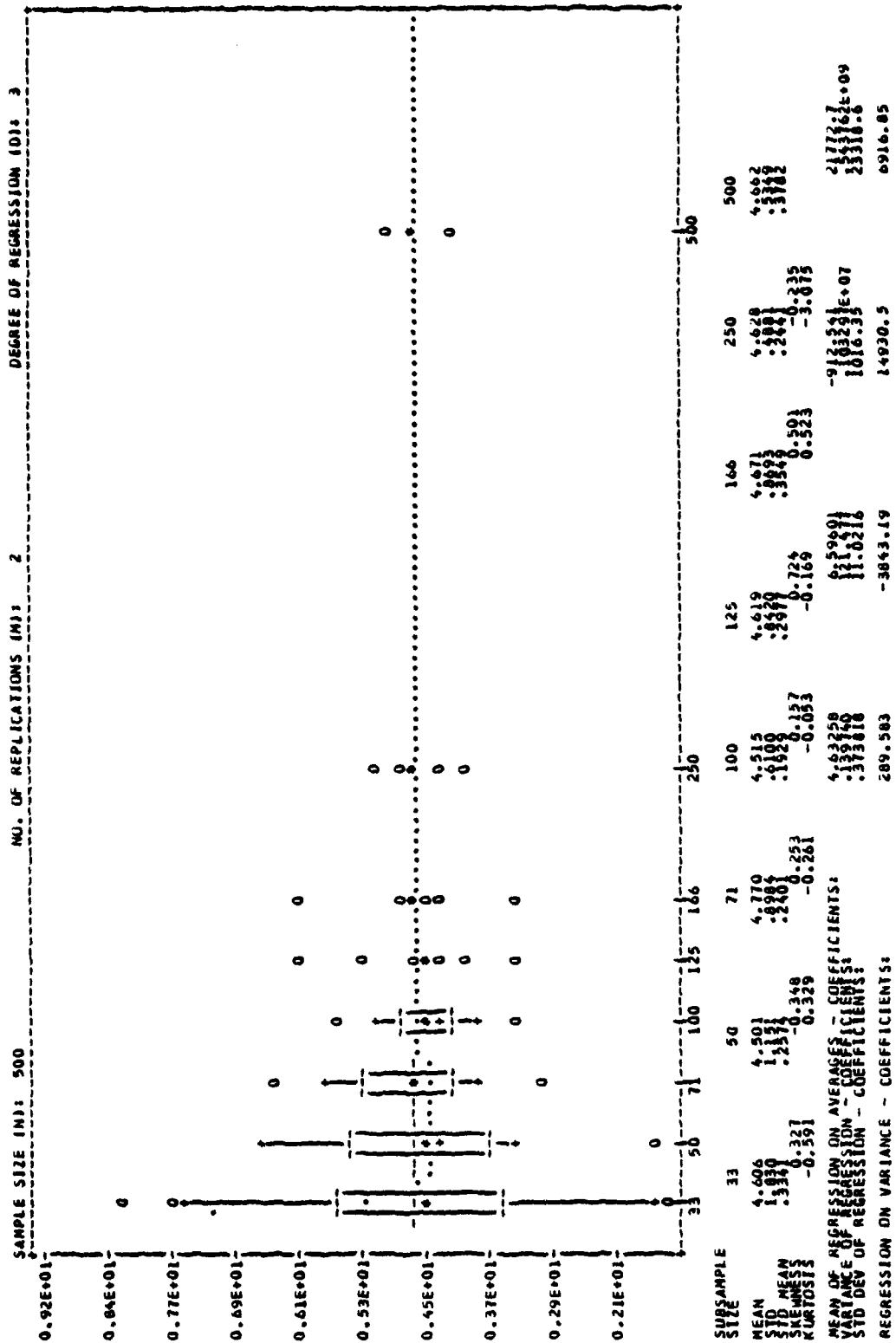
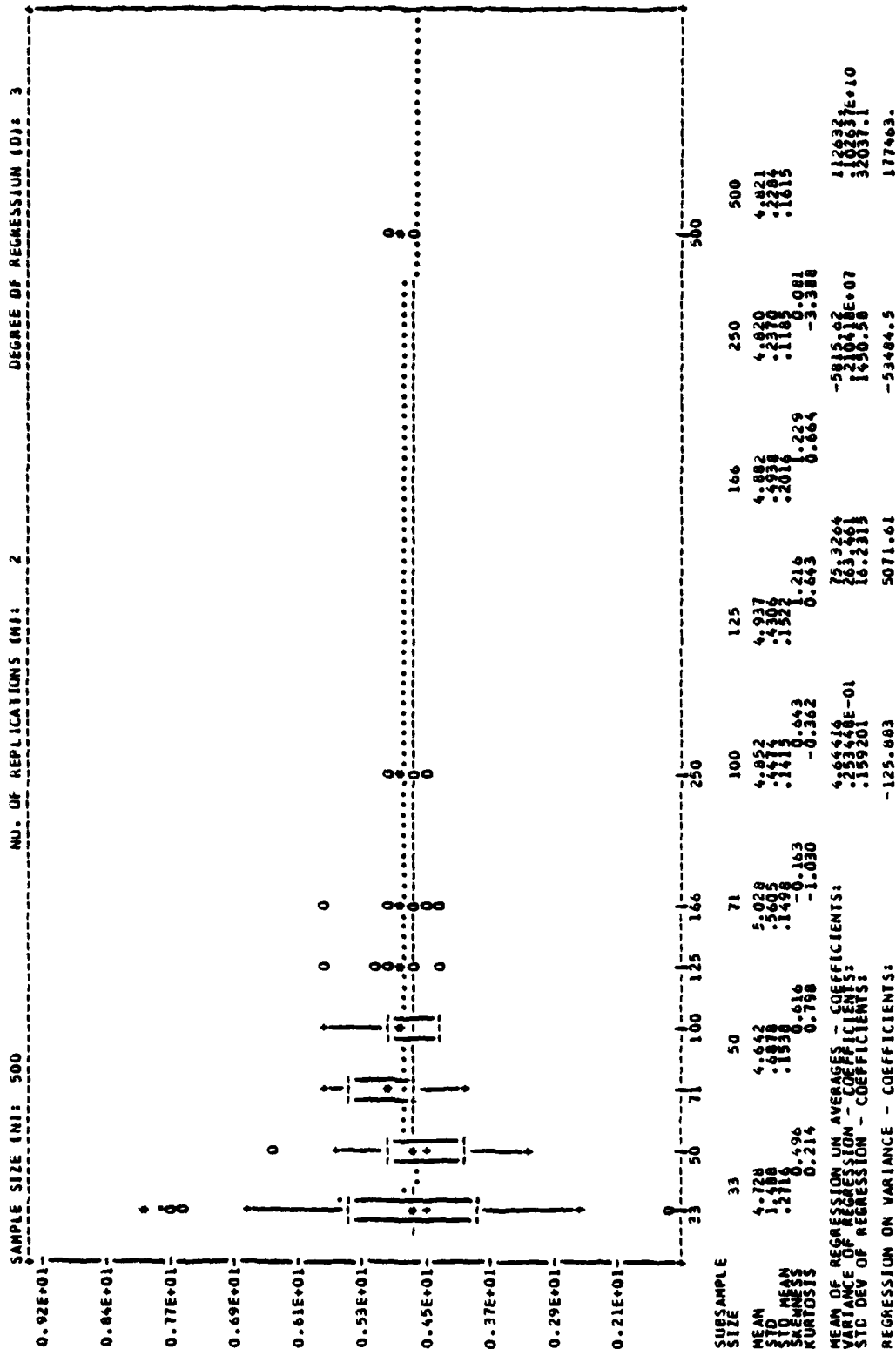


Figure 3c



ESTIMATOR: FOUR FOLD JACKKNIFED MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 3d

2. Case 1, Example 2: Small Subsample Sizes and Large Replications (K=5.0)

It was pointed out in Example 1 of this case that the small number of replications left some doubt as to the accuracy of the estimates of the bias coefficients as well as the normality of the estimates. Figures 4a to 4d (corresponding to Figures 3a through 3d) are for increased replications ( $M=100$ ). This has the effect of increasing the population of estimates for each subsample size ( $n_k$ ,  $k=1, \dots, K$ ) and the number of estimates of the bias coefficients ( $\beta_1, \beta_2, \dots$ ) but holds the subsample size at which the estimators are examined small as before. On first observing the graphs, it becomes apparent from the box plots of the two smallest subsample sizes ( $n_1=33$  and  $n_2=50$ ) that the estimates are nonnormally distributed (Figures 4a & 4b). The jackknifed estimators remove some of this nonnormality (Figures 4c & 4c), especially for the jackknifed moment estimator. In effect, outliers appear now at both high and low values but considerable skewness still remains.

Here it is apparent by the statistical output that the MLE estimator has significantly less variance in its population of estimates than does the moment estimator (Table 4).

Table 4: Standard deviations of the estimates of the shape parameter using the moment and MLE estimators for varying sample sizes (replications = 100)

SAMPLE SIZE	<u>Standard Deviation</u>	
	MOMENT ESTIMATOR	MLE ESTIMATOR
33	1.486	1.396
50	1.175	1.075
71	.9818	.8549
100	.8008	.6636
125	.7242	.5918
166	.6499	.5145
250	.5082	.4036
500	.3417	.2730

The increased symmetry and normality by the jackknifed moment estimator is graphically visible if Figures 4a & 4c are compared. The change in symmetry or normality of the estimates of the jackknifed MLE estimator was not discernable in the graphics (Figures 3b & 3d) or skewness coefficients (Table 3) for Example 1. However, for increased replications, (M=100) if the box plots of Figures 4b & 4d are compared it appears that the population of MLE estimates are more normally distributed (outliers at both ends) after jackknifing. The movement toward symmetry is more obvious by comparing the skewness coefficients of the MLE estimates and the jackknifed MLE estimates.

The variance reduction, or lack of it, for each estimator when it is jackknifed can be determined by comparing the computed variances that result from a given subsample size and the variance coefficients,  $\alpha_1$ ,  $\alpha_2$ , etc. If, for example, the variance coefficients ( $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ , and  $\alpha_4$ ) for each of the estimators (Figures 4a through 4d) are used to determine the variance for a subsample size of 500, the variances for each estimator would be:

$$\text{VAR(MOMENT)} = \frac{58.37}{500} + \frac{174.72}{500^{1.5}} - \frac{1500.13}{500^2} + \frac{5446.25}{500^{2.5}} = .1273$$

$$\text{VAR(MLE)} = \frac{67.52}{500} - \frac{961.91}{500^{1.5}} + \frac{10183.5}{500^2} - \frac{27503.3}{500^{2.5}} = .0848$$

$$\text{VAR(JNK MNT)} = \frac{79.20}{500} - \frac{457.44}{500^{1.5}} + \frac{5216.8}{500^2} - \frac{14533.1}{500^{2.5}} = .1357$$

$$\text{VAR(JNK MLE)} = \frac{62.66}{500} - \frac{825.2}{500^{1.5}} + \frac{8879.79}{500^2} - \frac{24484.2}{500^{2.5}} = .0827$$

These computations show a lack of change in the variance when the estimators are jackknifed and agree with the statistical and graphical conclusions drawn earlier.

The bias in the MLE and moment estimators is difficult to compare graphically (Figures 4a & 4b) since the unbiased estimate approaches the estimated expected value so quickly and because of the compacting caused by the range of the vertical scale. For this reason reduced graphics are produced (Figures 5a through 5d), which the reader will recall

suppresses the outliers. Since the scale is reduced from a range of 15.28 (14.00 - (-1.28)) to 8.10 (9.14 - 1.04) or almost 50%, the evolution of the bias as subsample size increases becomes clearer, as Figures 5a & 5b demonstrate in comparison to Figures 4a & 4b. These graphs indicate that, for at least small subsample sizes, the moment estimator is a more biased estimator than the MLE estimator.

Reduced graphics are not of much use for the jackknifed estimators (Figures 5c & 5d) since the bias is almost completely removed by jackknifing. They are useful only in the sense that they are on the same scale as their nonjackknifed forms and a comparison of bias between estimators can only be done graphically if the graphics of estimators are on the same scale. This, however, was already accomplished from Figures 4a through 4d.

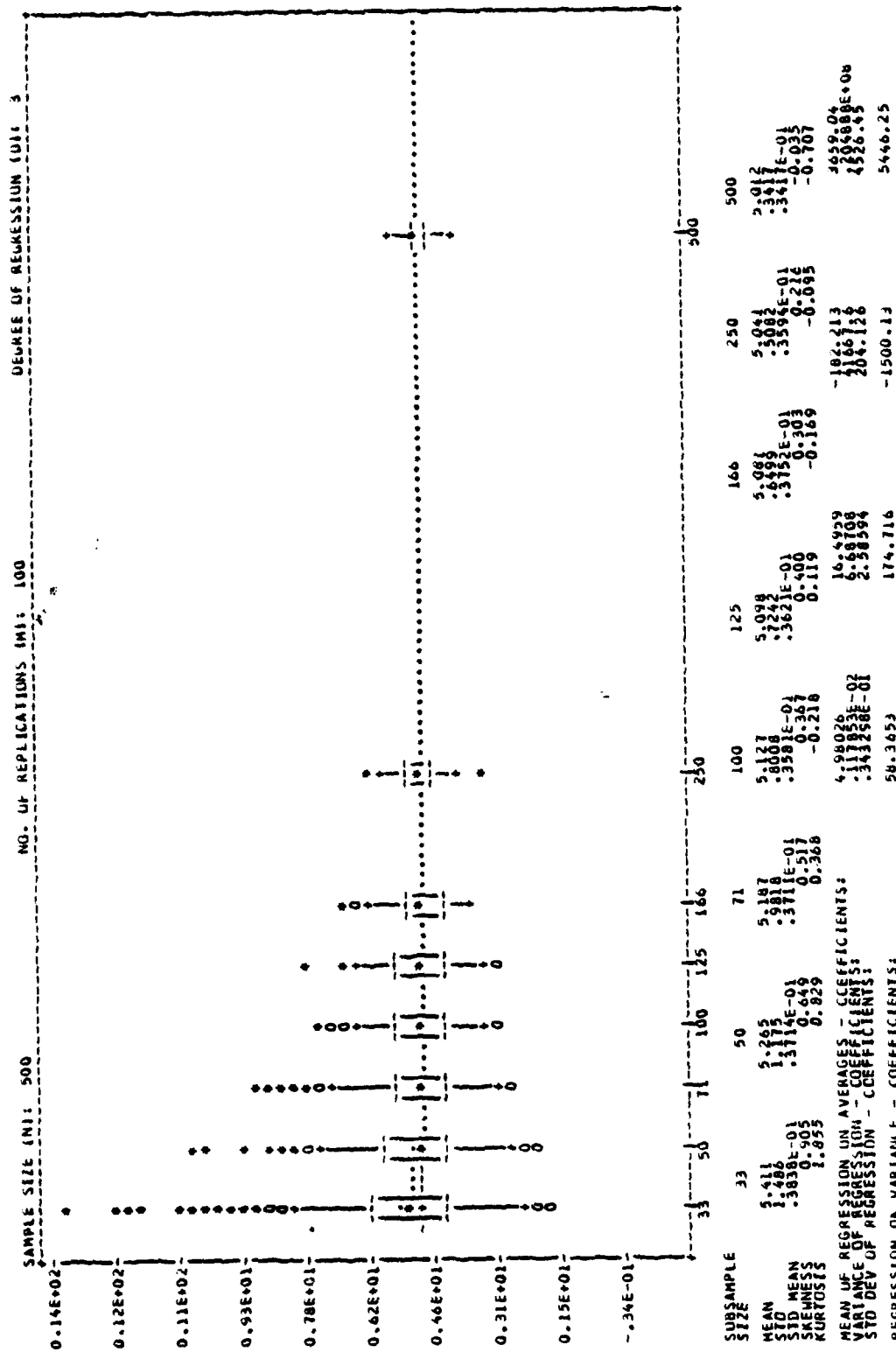
As was done for the variance, the amount of bias for a given subsample size can be estimated by dividing the coefficients,  $\beta_1$ ,  $\beta_2$ , etc., by the sample size squared, cubed, etc., as appropriate and summing (See Chapter II "USAGE" for the mathematical formula). The computations will not be carried out here as in the case of the variance but it is useful to compare the  $\beta_1$  coefficients of the estimators since these coefficients are likely to contribute the most significant amount of bias (they are divided by the smallest denominator). In the case of the moment estimator,  $\beta_1$  is

16.50 and for the MLE estimator it is 14.77. This suspected greater bias by the moment estimator can be seen visually from the graphics. The smaller coefficients for  $\beta_1$  of the jackknifed moment and MLE estimators are 1.47 and 4.09 respectively. This gives an indication of how much less the bias becomes when the estimators are jackknifed. This kind of comparison of  $\beta_1$  coefficients should be done with caution, though, since subsequent coefficients ( $\beta_2, \beta_3, \dots$ ) may be so large that they contribute more to the bias than the  $\beta_1$  coefficient. It should also be pointed out that these  $\beta_1$  coefficients should be zero because, as mentioned earlier, the jackknife removes the  $\beta_1(1/n)$  term. In fact, their estimated standard deviations are 4.00 and 5.67 respectively, so that they are within one standard deviation of zero.

As important as the size of these bias coefficients are their accuracy. Comparing the bias coefficients from Figures 5a through 5d in relation to their standard deviation with those of Figures 3a through 3d (Table 4) which were determined from only two replications, shows the advantage of large replication in estimating the bias of an estimator.

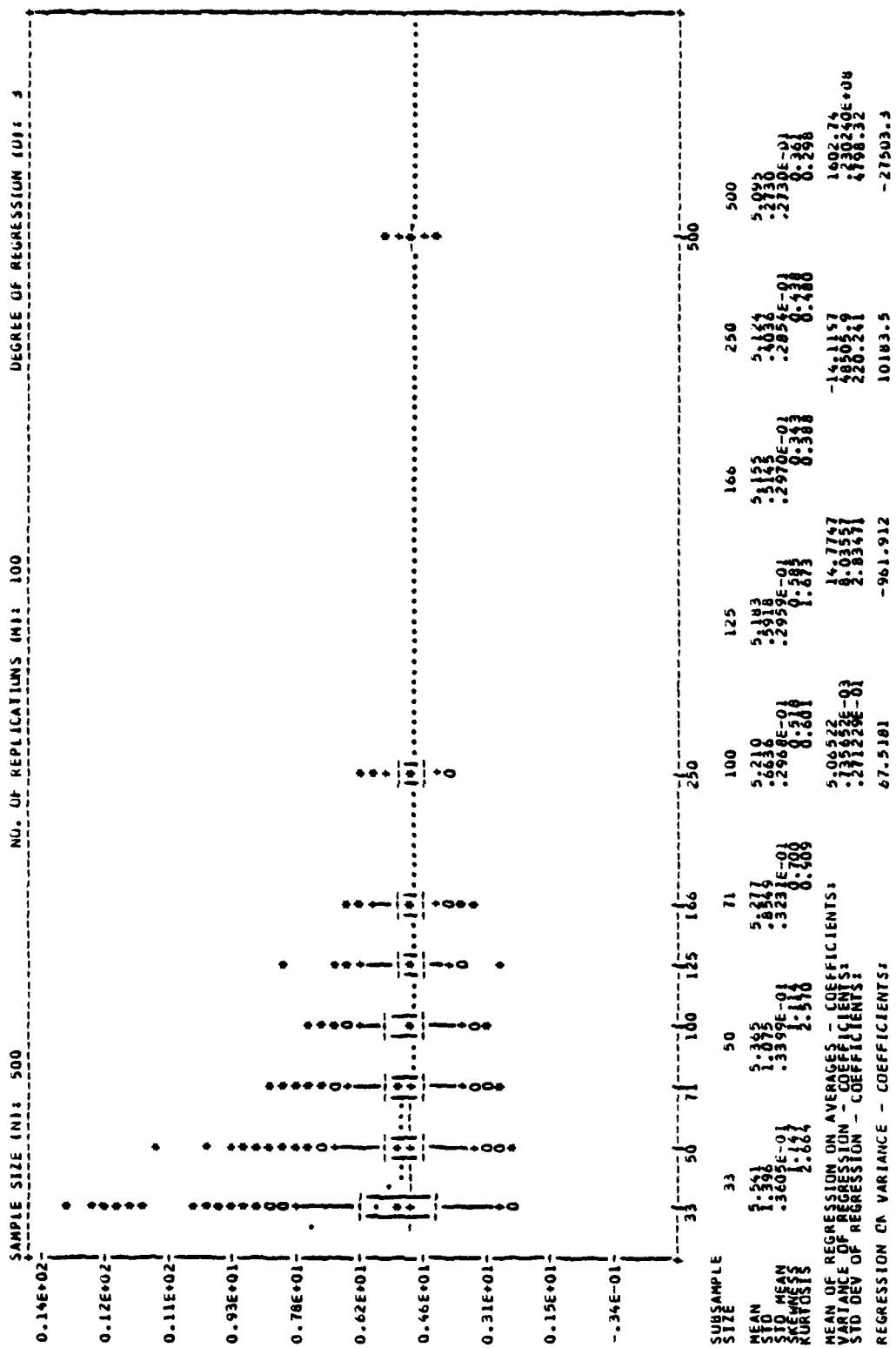
Table 5: Bias coefficients for each estimator's population of estimates and their standard deviations for small and large replications (M=2 and M=100)

ESTIMATOR	BIAS COEFF.	NUMBER OF REPLICATIONS	
		TWO (2)	ONE HUNDRED (100)
MOMENT	$\beta_1$	7.40 (13.5)	16.49 (2.59)
	$\beta_2$	183.1 (1105.5)	-182.2 (204.1)
	$\beta_3$	1947.9 (21284.5)	3659.04 (4526.5)
MLE	$\beta_1$	32.65 (19.1)	14.77 (2.83)
	$\beta_2$	-1998.7 (1031.2)	-14.12 (220.2)
	$\beta_3$	47308.0 (18073.1)	1602.74 (4798.3)
JNK MNT	$\beta_1$	6.59 (11.0)	1.47 (4.01)
	$\beta_2$	-912.54 (1016.4)	9.75 (335.5)
	$\beta_3$	21772.7 (23318.6)	-1369.26 (7411.7)
JNK MLE	$\beta_1$	75.33 (16.23)	4.09 (5.67)
	$\beta_2$	-5815.6 (1450.6)	-203.3 (436.02)
	$\beta_3$	112632.0 (32037.1)	793.11 (9243.2)



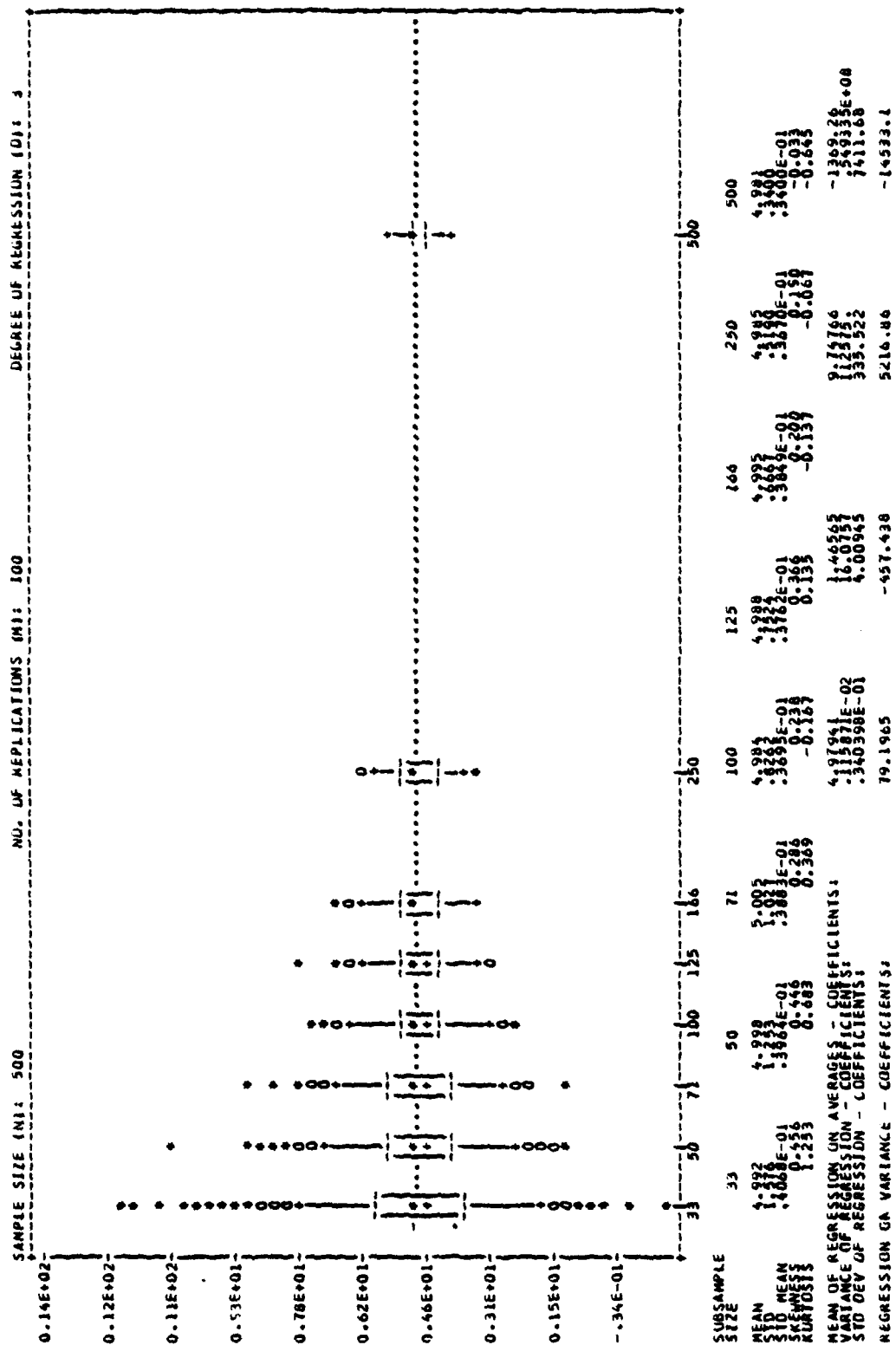
VERTICAL SCALE: YMIN = -1.2812  
 YMAX = 13.9984  
 ESTIMATOR: MOMENT ESTIMATOR (RECIPROCAL OF SQUARED COEFFICIENT OF VARIATION) OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 4a

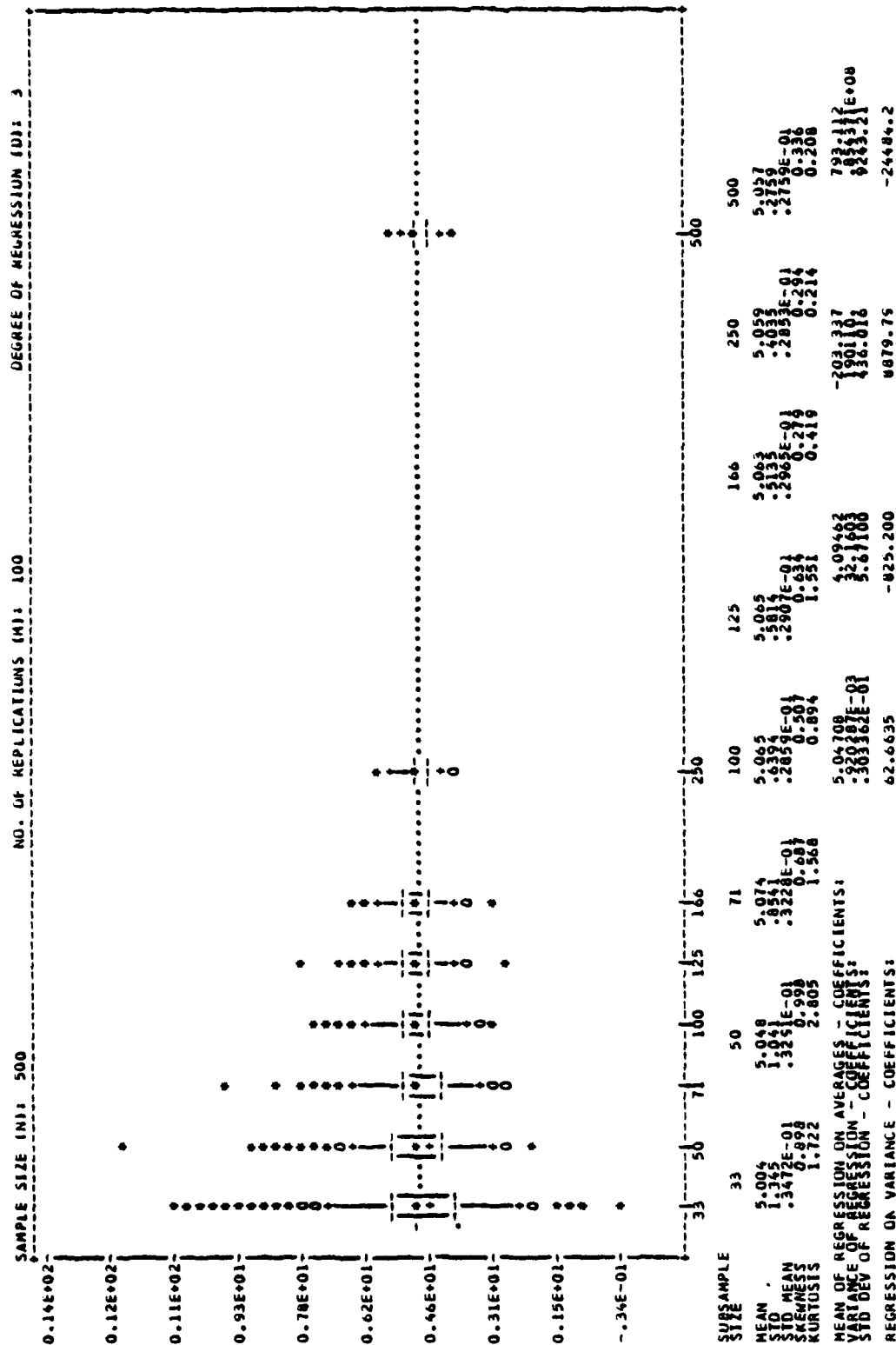


VERTICAL SCALE: YMIN = -1.2812  
YMAX = 13.9984  
ESTIMATOR: MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 4b



VERTICAL SCALE: MIN = 11.8012  
 ESTIMATOR: FOUR FOLD JACKKNIFED MOMENT ESTIMATOR OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.  
 Figure 4c



ESTIMATOR: FOUR FOLD JACKKNIFED MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 4d

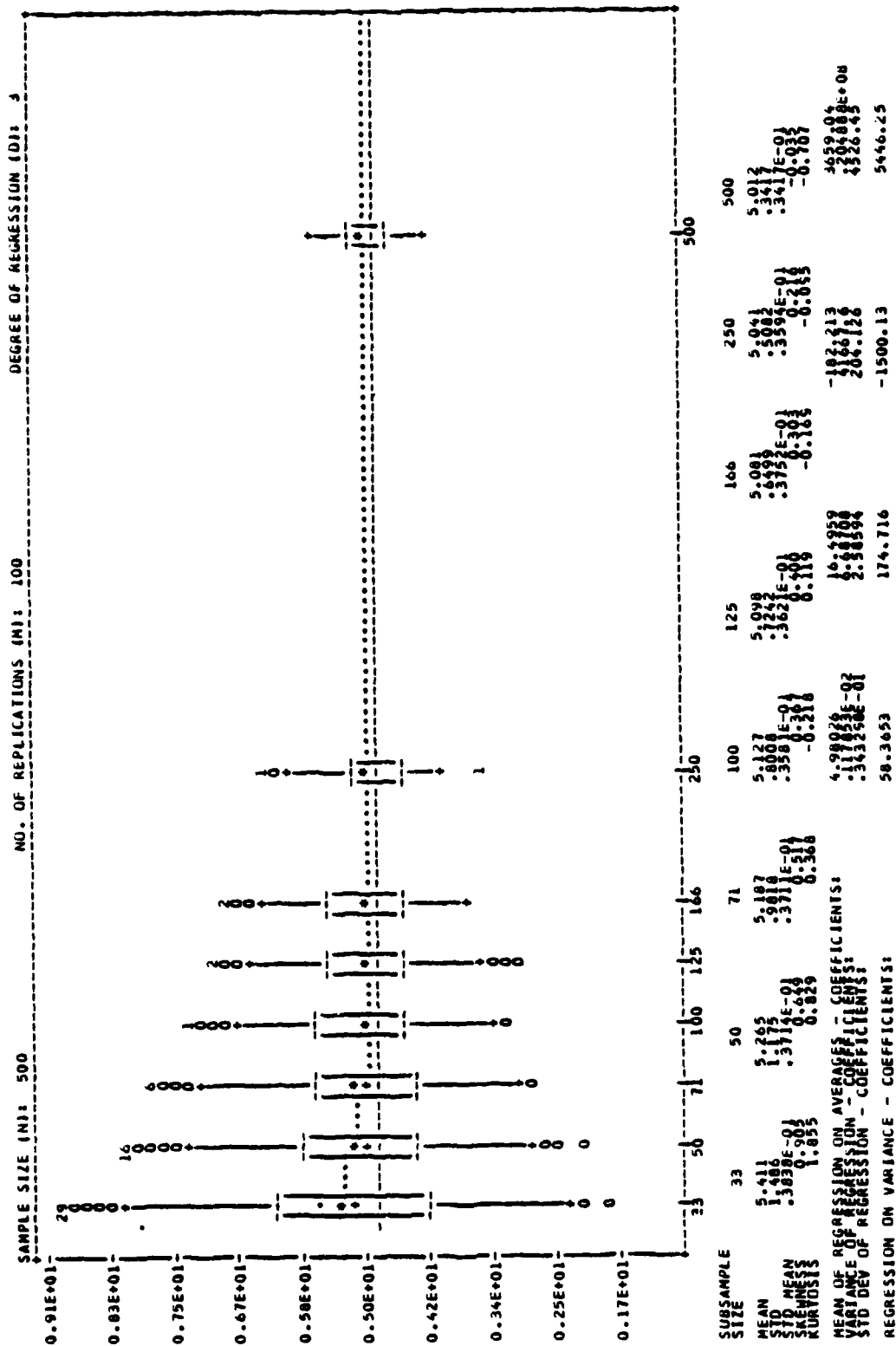
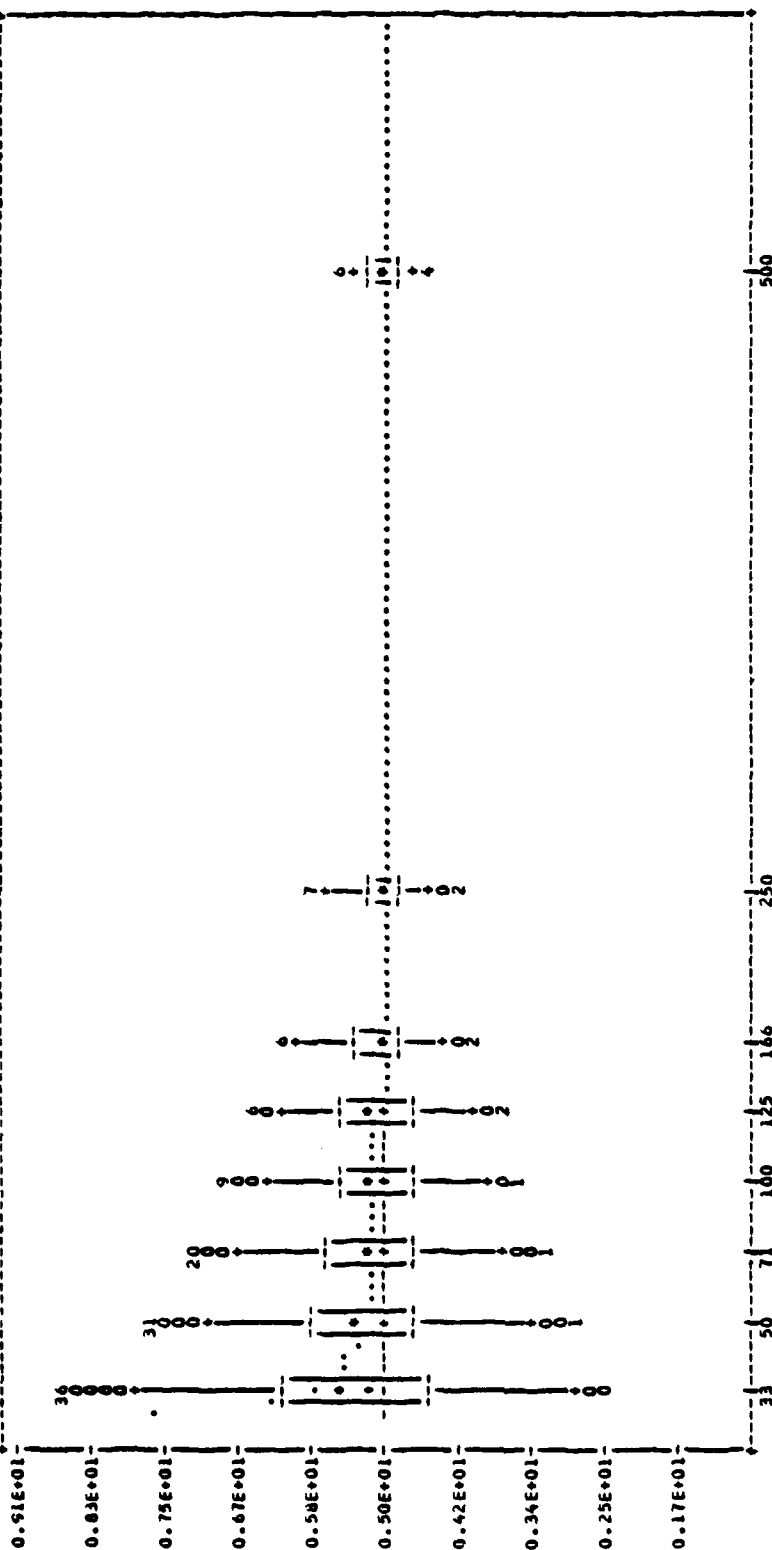


Figure 5a (REDUCED GRAPHICS)

SAMPLE SIZE (N): 500

NU. OF REPLICATIONS (M): 100

DEGREE OF REGRESSION (D): 3



SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	1.541	1.393	1.377	1.210	1.183	1.155	1.134	1.093
STD. MEAN	1.386	1.073	1.037	0.830	0.803	0.775	0.734	0.693
VARIANCE	.360E-01	.339E-01	.323E-01	.296E-01	.295E-01	.297E-01	.285E-01	.273E-01
STD. DEV.	1.147	1.116	1.100	0.930	0.923	0.930	0.840	0.781
KURTOSIS	2.884	2.310	0.509	0.601	1.673	0.388	0.480	0.298
MEAN OF REGRESSION UN AVERAGES - COEFFICIENTS:				5.06522	14.7747		-14.1157	1602.74
VARIANCE OF REGRESSION UN AVERAGES - COEFFICIENTS:				.73545E-03	6.03457		40304.9	230240E+08
STD. DEV. OF REGRESSION UN AVERAGES - COEFFICIENTS:				.271229E-01	2.83471		220.241	4898.32
REGRESSION ON VARIANCE - COEFFICIENTS:				67.5181	-961.912		10183.5	-27503.3
VERTICAL SCALE: YMIN =								
YMAX =								
ESTIMATOR: MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.								

Figure 5b (REDUCED GRAPHICS)

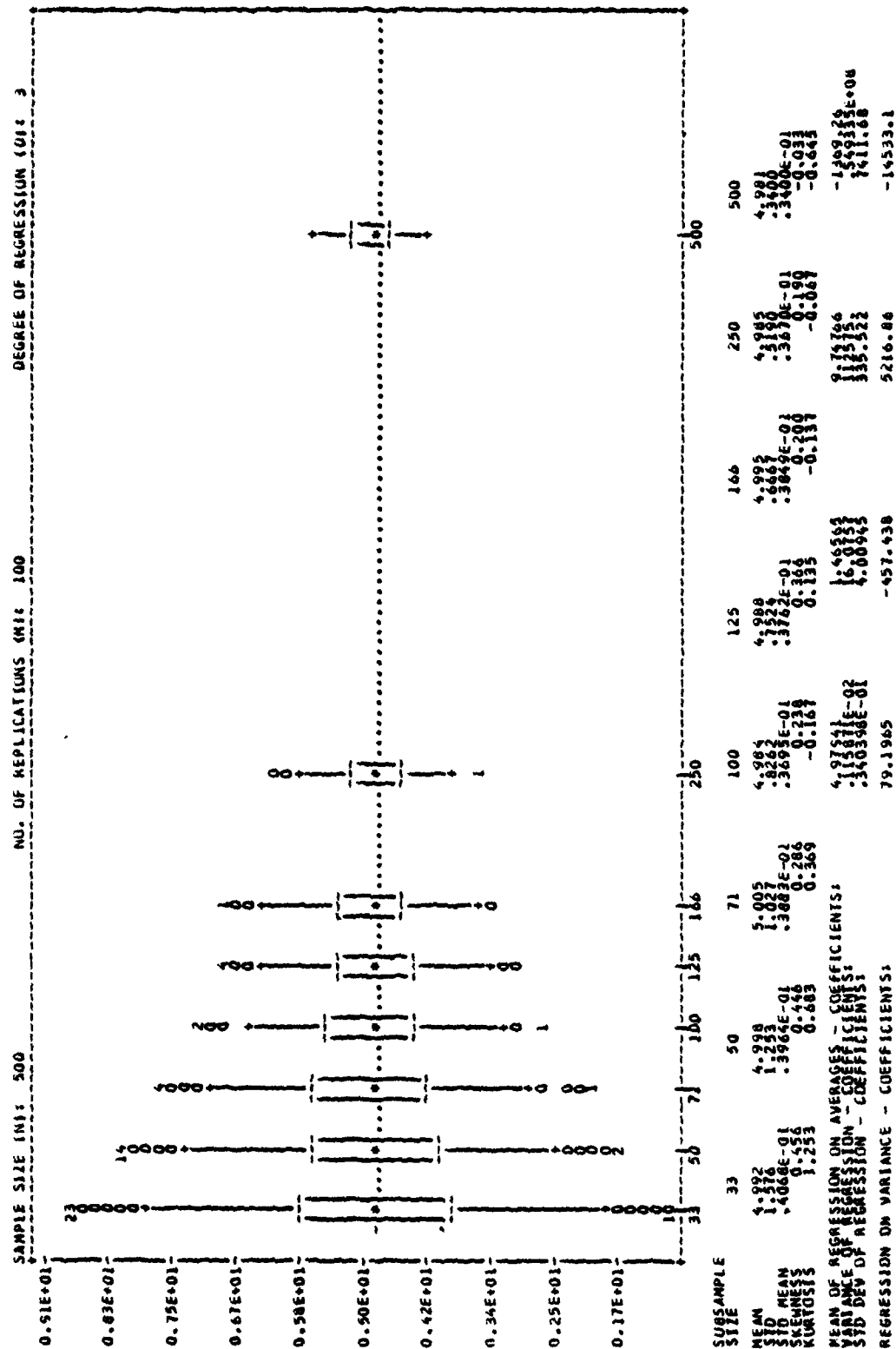


Figure 5c (REDUCED GRAPHICS)

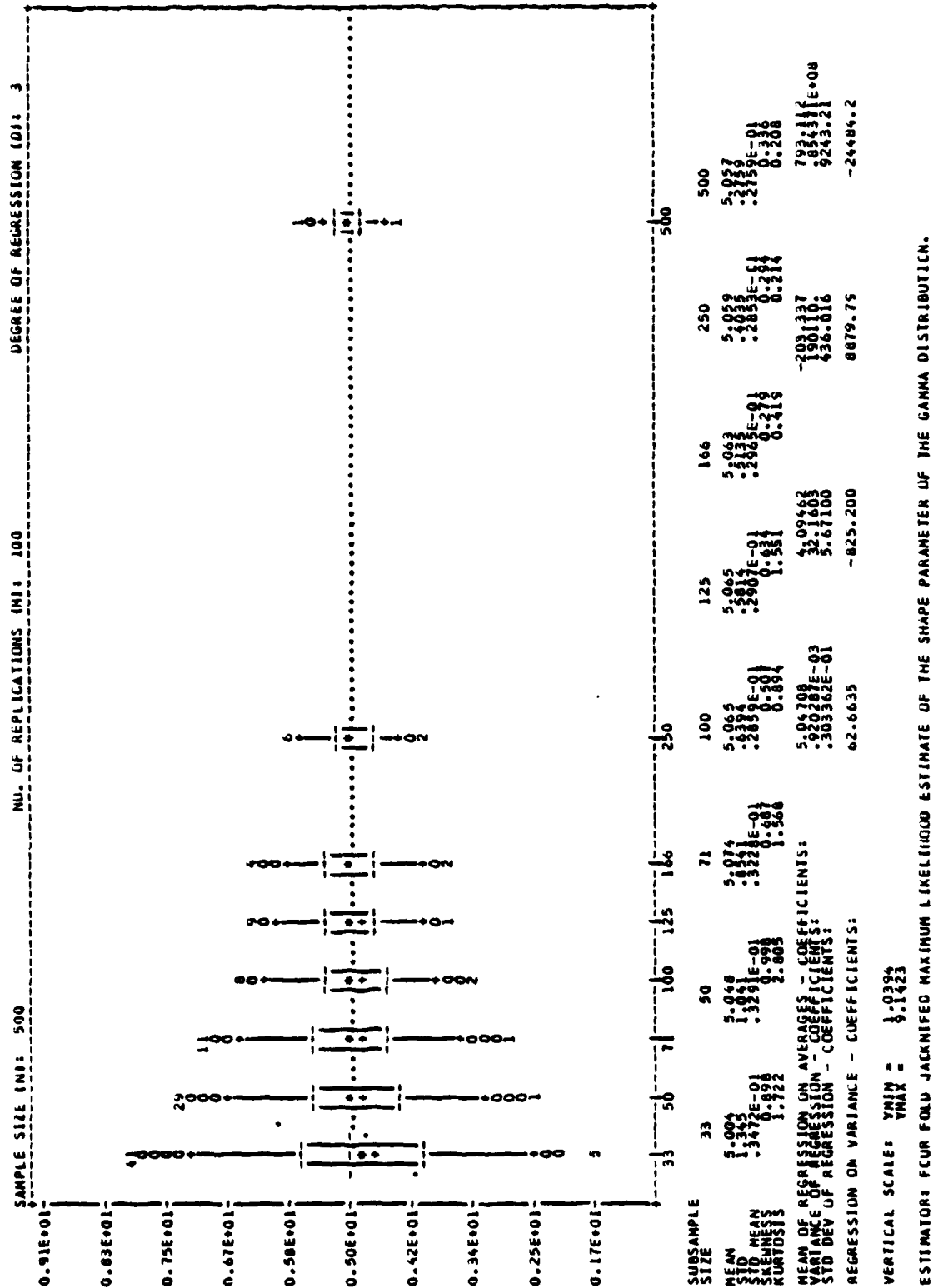


Figure 5d (REDUCED GRAPHICS)

3. Case 1, Example 3: Large Subsample Sizes and Large Replications ( $K=5.0$ )

Figures 4a through 4d showed the necessity of increasing subsample sizes in order to examine the normality of the estimates. Figures 6a through 6d are for increased subsample sizes (100 to 2000) with a large number of replications ( $M=25$ ). The box plots in these figures show the increase in normality with increasing subsample size  $n_k$  when they are compared with small subsample sizes (Figures 4a through 4d). The reduction in variability is also obvious by noticing the lack of outliers in the box plots and the smaller range of the vertical scale for larger sample sizes. The standard deviation for each of the estimators for small subsample sizes ( $n_g=500$ , Figures 4a through 4d) is compared to those for large subsample sizes ( $n_g=2000$ , Figures 6a through 6d), Table 6. It can be seen that the variance is reduced for larger subsample sizes by the appropriate factor of two. However, the degrees of freedom in the standard deviation of the estimators are 25 at  $n_g = 2000$  and 100 at  $n_g = 500$ . Again it is clear from these figures that jackknifing causes no lack of precision in either the MLE or moment estimators.

The bias in the moment and MLE estimators is difficult to compare from the graphics because of the range of the vertical scale. For this reason reduced graphics (Figures 7a & 7b) for these two estimators are produced. Reduced graphics for the jackknifed estimators would not be very useful since it

has been shown in previous examples that very little bias exists. For this reason they are omitted.

Table 6: Standard deviations of the population of estimates of the shape parameter for each estimator for subsample sizes 500 and 2000

ESTIMATOR	SAMPLE SIZES	
	$n_8 = 500$	$n_6 = 2000$
MOMENT	.3417	.1592
MLE	.2730	.0910
JACKKNIFED MOMENT	.3400	.1597
JACKKNIFED MLE	.2759	.1090

A change seems to have occurred if a comparison is made of Figures 7a & 7b with those of Figures 5a & 5b. The graphics seem to indicate that the MLE's unbiased estimates have shifted from being biased on the high side to the low side for subsample size  $n_1=100$  and possibly  $n_2=200$ . Also, it appears that the two estimators have about the same bias or that possibly the moment estimator has less bias. Looking further at the bias coefficients that determine these graphics, it can be seen that the coefficients in Figures 5a and 5b are estimated with more precision than those of Figures 7a and 7b. One may be tempted to assume that the MLE estimator is more biased for larger sample sizes. But this is not likely and the change in the bias structure for Figure 7b can be

attributed to the difference in the values of the bias coefficients ( $\beta_1, \beta_2, \dots$ ) compared to those of Figure 5b. These coefficients should be the same but the coefficients of Figure 5b were computed from 100 estimates (100 regressions) and those from Figure 7b only 25. The precision can be expected to be better in the first case. This is obviously the case if the standard deviation of the coefficients are compared, and more confidence can be placed in the bias structure of Figure 5b.

SAMPLE SIZE (N): 2000

NU. OF REPLICATIONS (M): 25

DEGREE OF REGRESSION (D): 3

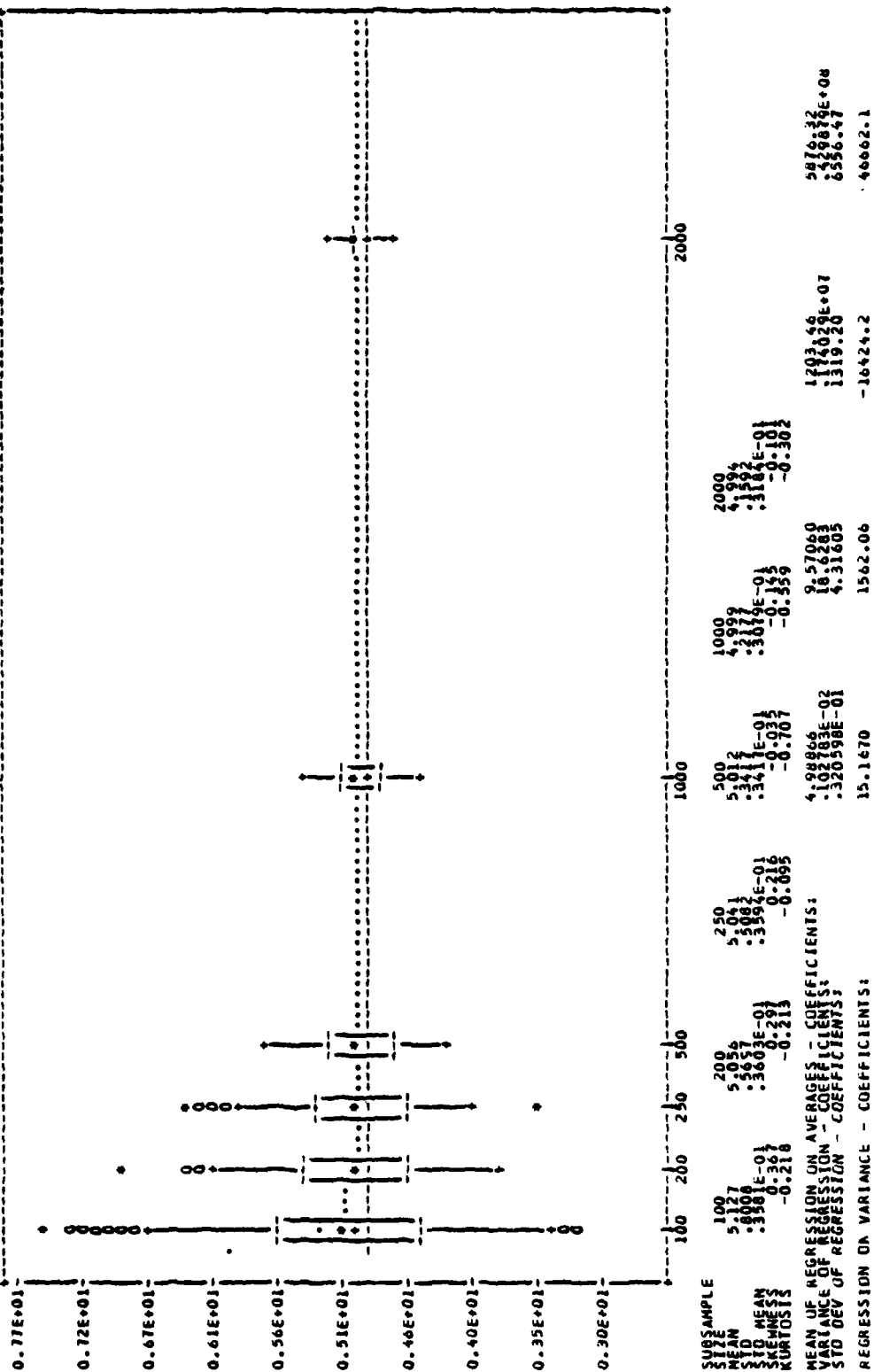
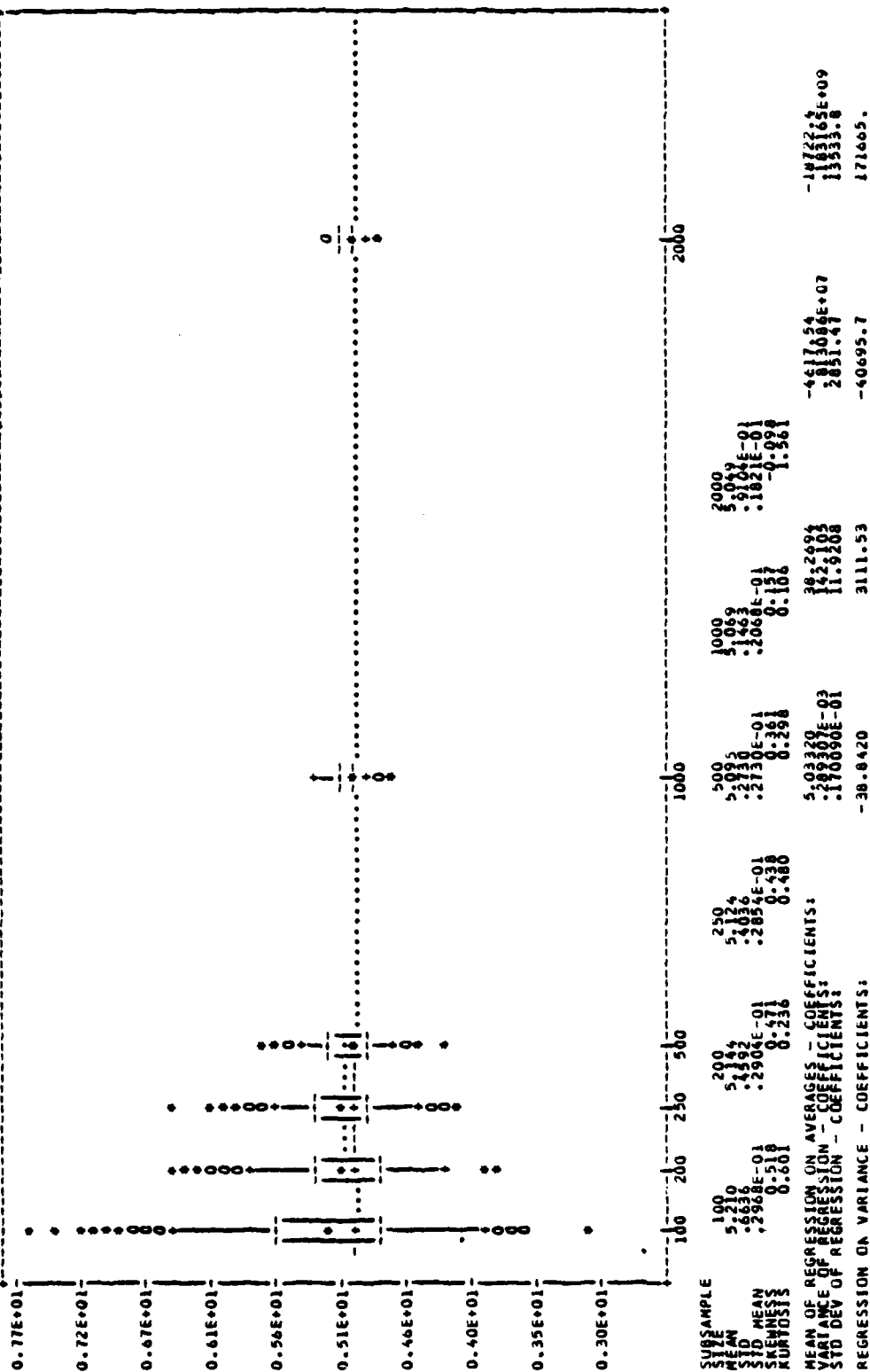


Figure 6a

SAMPLE SIZE (N): 2000

NUM. OF REPLICATIONS (M): 25

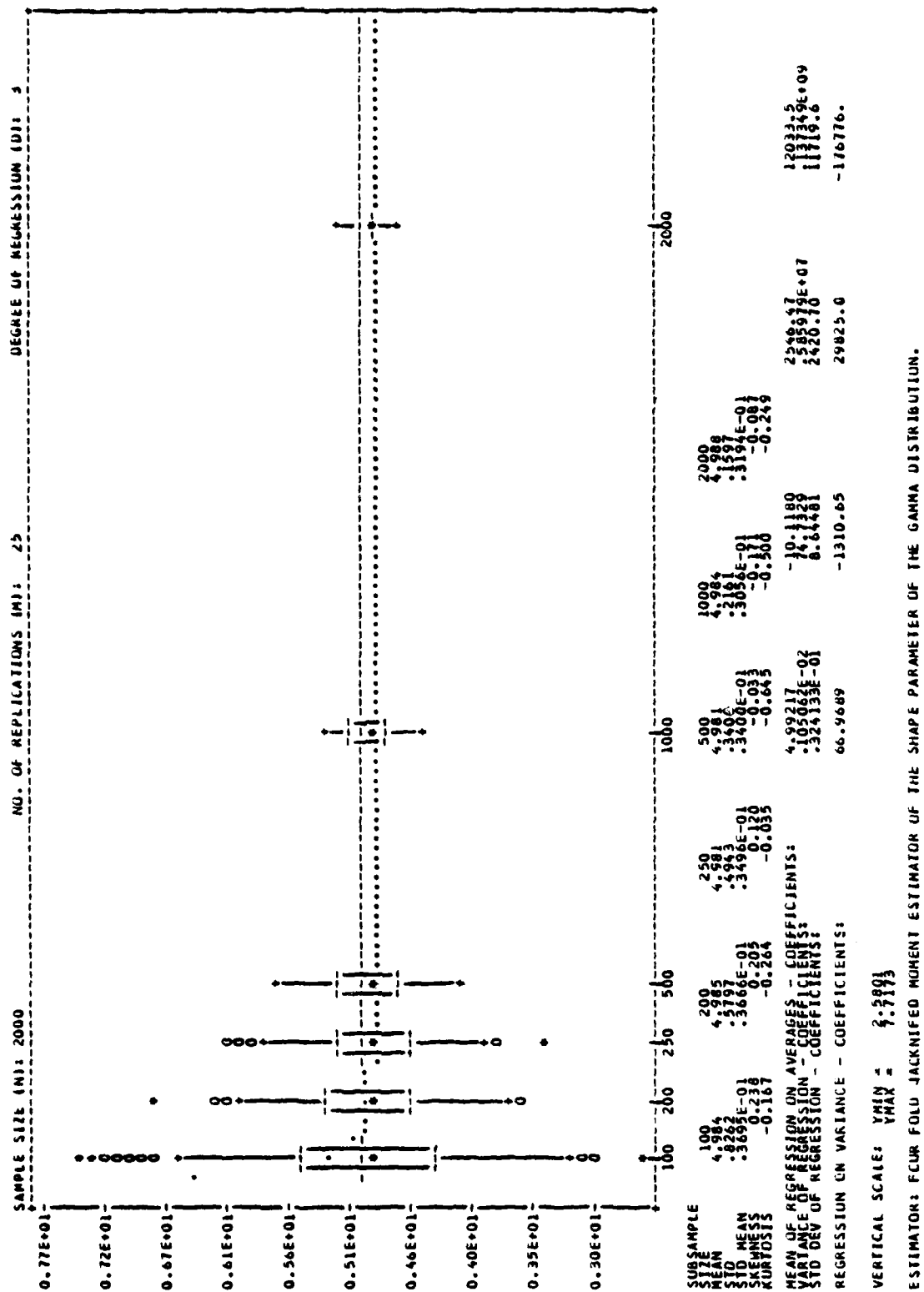
DEGREE OF REGRESSION (D): 4



VERTICAL SCALE: YMIN = 2.5001  
YMAX = 7.7173

ESTIMATOR: MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 6b





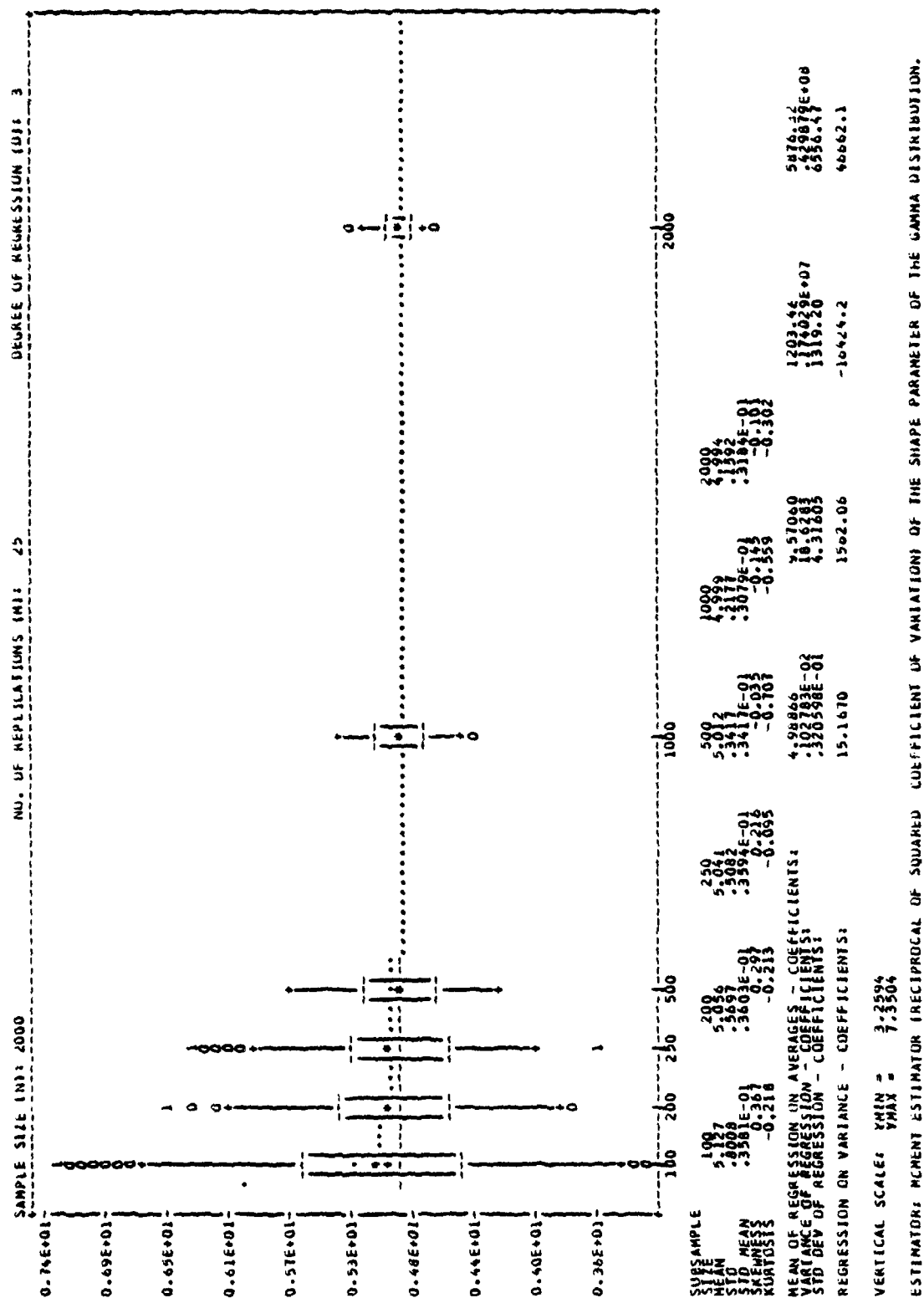


Figure 7a (REDUCED GRAPHICS)

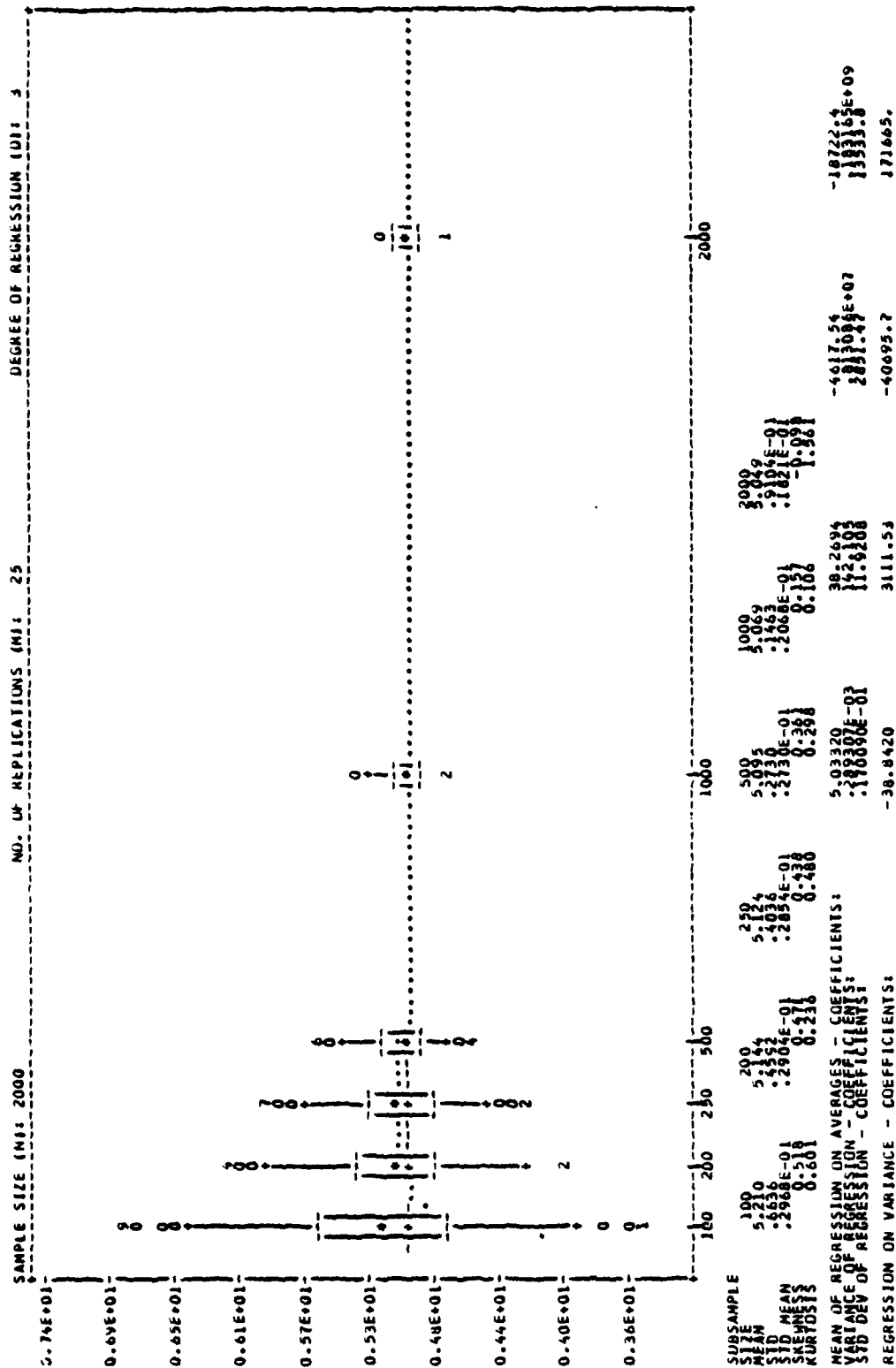


Figure 7b (REDUCED GRAPHICS)

B. EXAMINING THE GAMMA DISTRIBUTION FOR THE SHAPE PARAMETER  
( $K=0.2$ )

1. Case 2, Example 1: Small Subsample Sizes and  
Small Replications ( $K=0.2$ )

In the preceding case for  $K=5.0$  four estimators of the shape parameter of the gamma distribution were evaluated for varying subsample sizes and replications. This evaluation is repeated here with the shape parameter changed from  $K=5.0$  to  $K=0.2$ . This second set of comparisons is not only useful in demonstrating how the RAGE package shows the behavior of different estimators in regards to bias and variance, but also how these estimators may differ when the parameter being evaluated takes on different values.

Figures 8a through 8d are graphical results of the four estimators of the shape parameter,  $K=0.2$ . For these small subsample sizes and small population of estimates it is apparent from the graphics that the MLE estimator (Figure 8b) has tighter box plots than does the moment estimator (Figure 8a). This shows graphically that the variance for the MLE estimator is less than that of the moment estimator. This was not obvious when the shape parameter was 5.0. In Figures 3a & 3b of Case 1, Example 1, it was difficult to distinguish any difference in the box plots. This comparison shows how the MLE estimator is much better at reducing the variance when the actual shape parameter is as small as 0.2. Or, alternatively, it may be an indication of how much worse

the moment estimator is for estimating shape parameters that are small.

Figures 8c & 8d are the jackknifed forms of the estimators. It is not at all clear from the graphs that the jackknife is affecting the variance of the estimates. This lack of change could be attributed to the small number of estimates which suggest that the number of replications should be increased to achieve a better graphical evaluation.

In Case 1, Examples 1, 2 and 3, the coefficients of variance ( $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  and  $\alpha_4$ ) were helpful in determining which estimators produced the least variance. A look at  $\alpha_1$  in Figures 8b and 8d shows that they are negative numbers. Since  $\alpha_1$  often contributes the most significant amount to the variance, there is some suspicion that a determination of the variance by the use of these coefficients could result in negative variances for the MLE estimator and its jackknifed form. If  $\alpha_2$  and  $\alpha_3$  are not large positive numbers (relatively) it is especially likely. In these two cases  $\alpha_2$  is a relatively large positive number and negative variances do not occur if the calculations are carried through. However, for lesser degrees of regression (i.e.  $D=2$ ) negative variances can occur. This is a result of the wide fluctuation of the variances of the small subsample sizes whose values are used as dependent variables in the regression which determine  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ . These variance coefficients are often

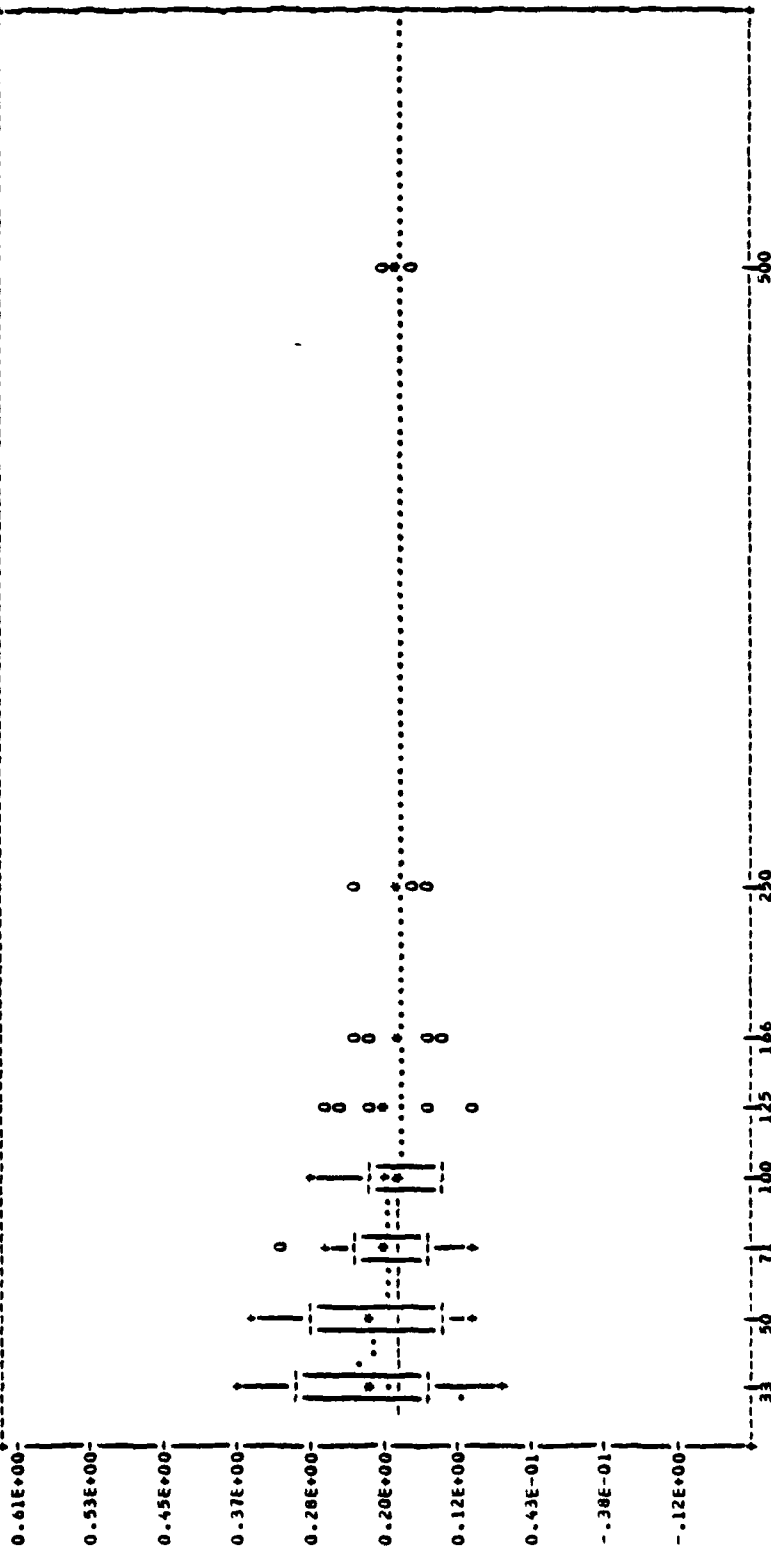
questionable when the variances of the population of estimates vary over a wide range and further suggest increasing the number of replications so that the variance coefficients can be more accurately determined.

The bias in the moment estimator (Figure 8a) appears to be considerably more than that of the MLE estimator (Figure 8b) when the graphs of figures are compared. This bias is further reduced when the jackknife is applied to the estimators (Figures 8c & 8d). A comparison of the bias is inadvisable at this point since the bias coefficients have large standard deviation values in comparison to their estimates. The  $\beta_1$  coefficient, for example, of the MLE estimator is .0864 and its standard deviation is .189. As mentioned in previous examples,  $\beta_1$  will normally contribute the most to the bias and since its standard deviation is larger than the estimate itself, its accuracy is highly questionable. Before any firm conclusions can be drawn concerning the bias of the estimators more estimates are required, which implies increasing the number of replications.

SAMPLE SIZE (N): 500

NO. OF REPLICATIONS (M): 2

DEGREE OF REGRESSION (D): 3



SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	.2445	.2207	.2099	.2030	.2045	.1951	.1921	.1944
STD	.0243E-01	.0290E-01	.0271E-01	.0231E-01	.0233E-01	.0254E-01	.0243E-01	.0243E-01
STD MEAN	.1505E-01	.1675E-01	.1542E-01	.1329E-01	.1887E-01	.1453E-01	.1673E-01	.1617E-01
SKEWNESS	-0.167	-0.064	0.278	0.411	-0.260	0.393	0.461	0.461
KURTOSIS	-1.090	-1.180	-0.483	-0.873	-1.451	-1.443	-0.946	-0.946
MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:				.1227E-03			.7105E-03	
VARIANCE OF REGRESSION - COEFFICIENTS:				.155073E-01			.8927E-03	
STD DEV OF REGRESSION - COEFFICIENTS:				.718084			.9427E-03	
REGRESSION ON VARIANCE - COEFFICIENTS:					-12.9665		117.237	
REGRESSION ON VARIANCE - COEFFICIENTS:								-1.561E-07
REGRESSION ON VARIANCE - COEFFICIENTS:								2503.12
REGRESSION ON VARIANCE - COEFFICIENTS:								-339.596

VERTICAL SCALE: YMIN = -0.1829  
YMAX = 0.6069

ESTIMATOR: MOMENT ESTIMATOR RECIPROCAL OF SQUARED COEFFICIENT OF VARIATION OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 8a



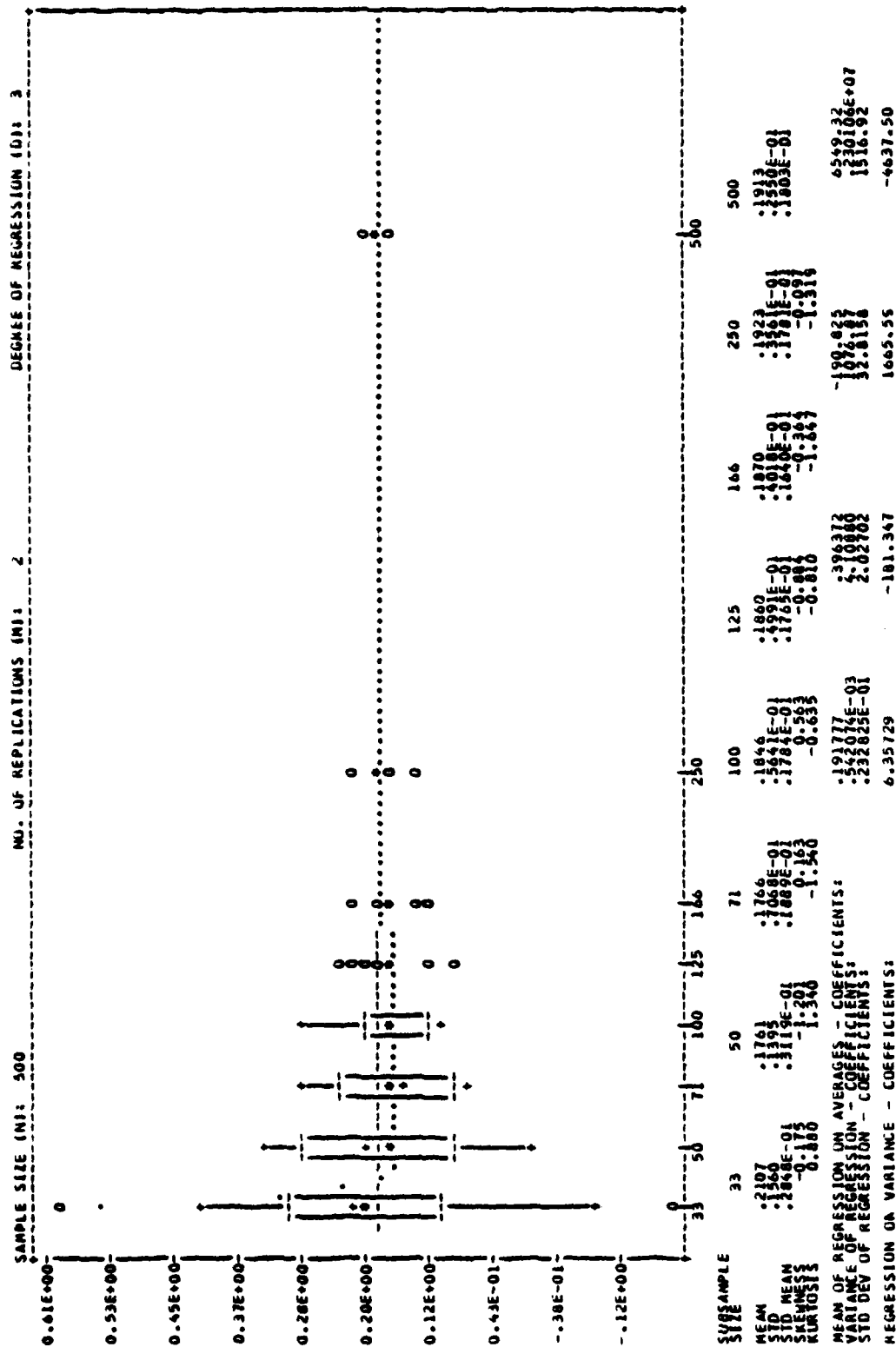


Figure 8c

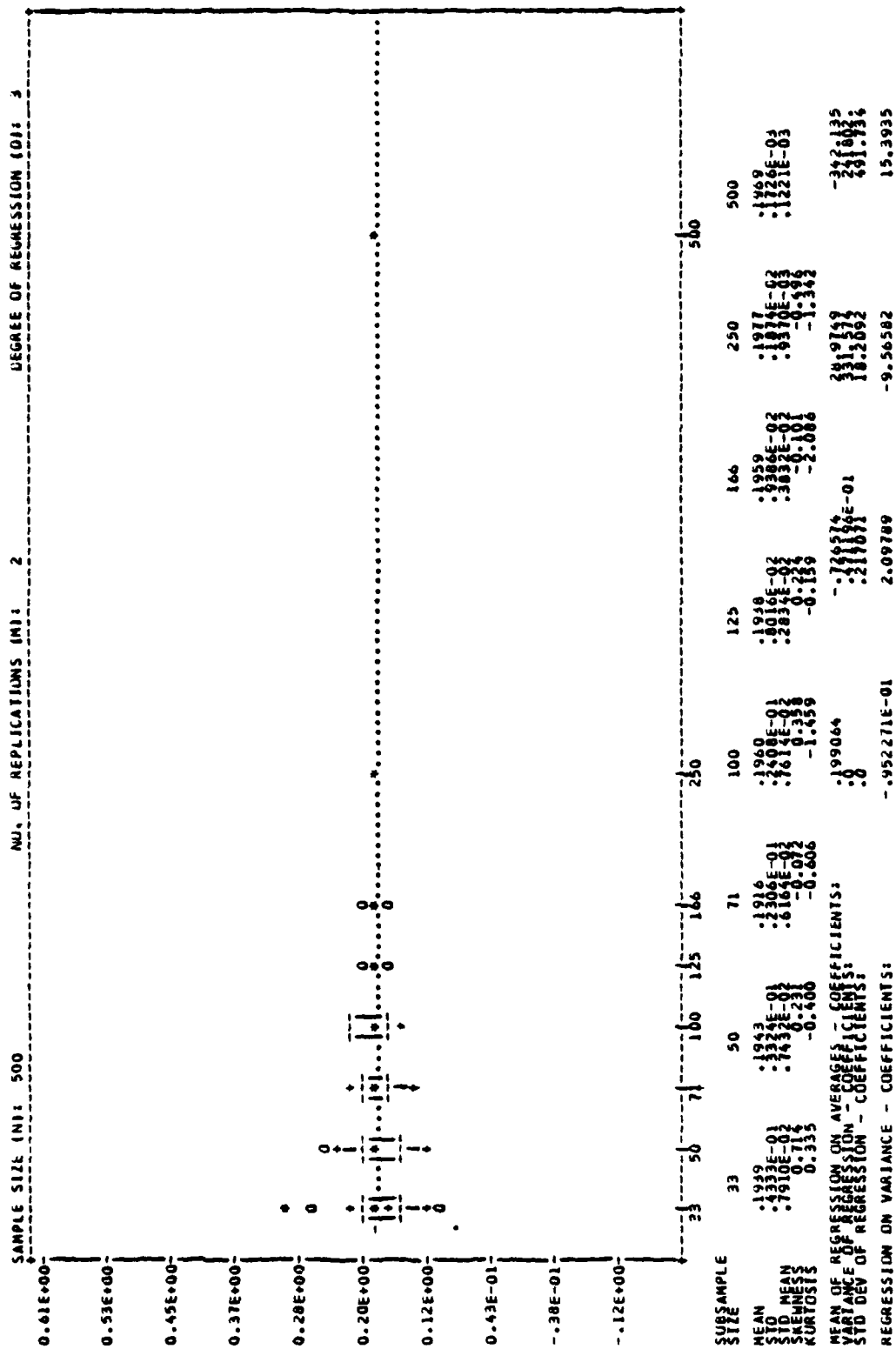


Figure 8d

2. Case 2, Example 2: Small Subsample Sizes and Large Replications ( $K=0.2$ )

Figures 9a through 9d are for small subsample sizes but large replications ( $M=100$ ). This increased number of estimates produces more definitive graphics for better evaluations of the bias and variance of the estimates produced by the estimators.

The graphs of the moment and MLE estimators (Figures 9a & 9b) show clearly as before that the MLE estimator has much less variance than does the moment estimator. Both, however, have positive skewness and high nonnormal estimates up to subsample size  $n_4=125$ . The jackknifed form of the moment estimator has about the same variance as its nonjackknifed form but is more symmetrically distributed. This can be seen in a comparison of the box plots of Figures 9a & 9c.

The distribution of the population of estimates for the jackknifed MLE estimator shows little if any difference graphically or statistically than the nonjackknifed estimates. This can be seen by comparing Figures 9b & 9d, which show little difference when the box plots are compared for various subsample sizes. For large values of the shape parameter, i.e.  $K=5.0$ , it could be seen from Figures 4b & 4d of Case 1, Example 1 that the jackknifed MLE estimator increased the symmetry of the estimates. This was obvious both in the graphics and by comparing the skewness coefficients. For  $K=0.2$  the only indication that the estimates

are more symmetrically distributed is the comparison of the skewness coefficients for small subsample sizes ( $n_1=33$ ,  $n_2=50$  and  $n_3=71$ ) of the MLE estimator and its jackknifed form (Figures 9b & 9d). The skewness appears to decrease slightly, but not discernably, for larger subsample sizes.

The amount of variance that was observed graphically can also be seen statistically by computing the variance of each estimator for designated subsample sizes using the variance coefficients  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  and  $\alpha_4$  of each estimator. These coefficients for each estimator will result in positive variances. This is not always the case when the coefficients are determined from subsample sizes with a small population of estimates. This example illustrates the increased confidence that can be put into the variance coefficients as the number of estimates (replications) increases.

The bias again appears to be substantially greater for the moment estimator than the MLE estimator for comparable subsample sizes. Here, because of the range of the vertical scale, reduced graphics are produced for the moment and MLE estimators to better compare the bias of these estimators by reducing the scale (Figures 10a and 10b). A comparison of Figures 10a and 10b clearly shows the increased bias caused by the moment estimator. It was shown in Example 1 of Case 1 that the jackknifed estimators have little bias remaining in the estimates. This appears to be the case for

$K=0.2$ , as indicated by the small bias coefficients of the jackknifed estimators (Figures 9c & 9d). For this reason, reduced graphics for the jackknifed estimators are omitted, since indications are they would show little more graphically than Figures 9c & 9d.

It is interesting to note that the bias of the moment estimator compared to that of the MLE estimator is not only greater, but greater for small values of the shape parameter (i.e.  $K=0.2$ ) than for large values (i.e.  $K=5.0$ ). This can be seen by comparing the bias in Figures 10a & 10b with that of Figures 5a & 5b. This indicates that the moment estimator is not only more biased than the MLE estimator but considerably more biased when the shape parameter being estimated is small.

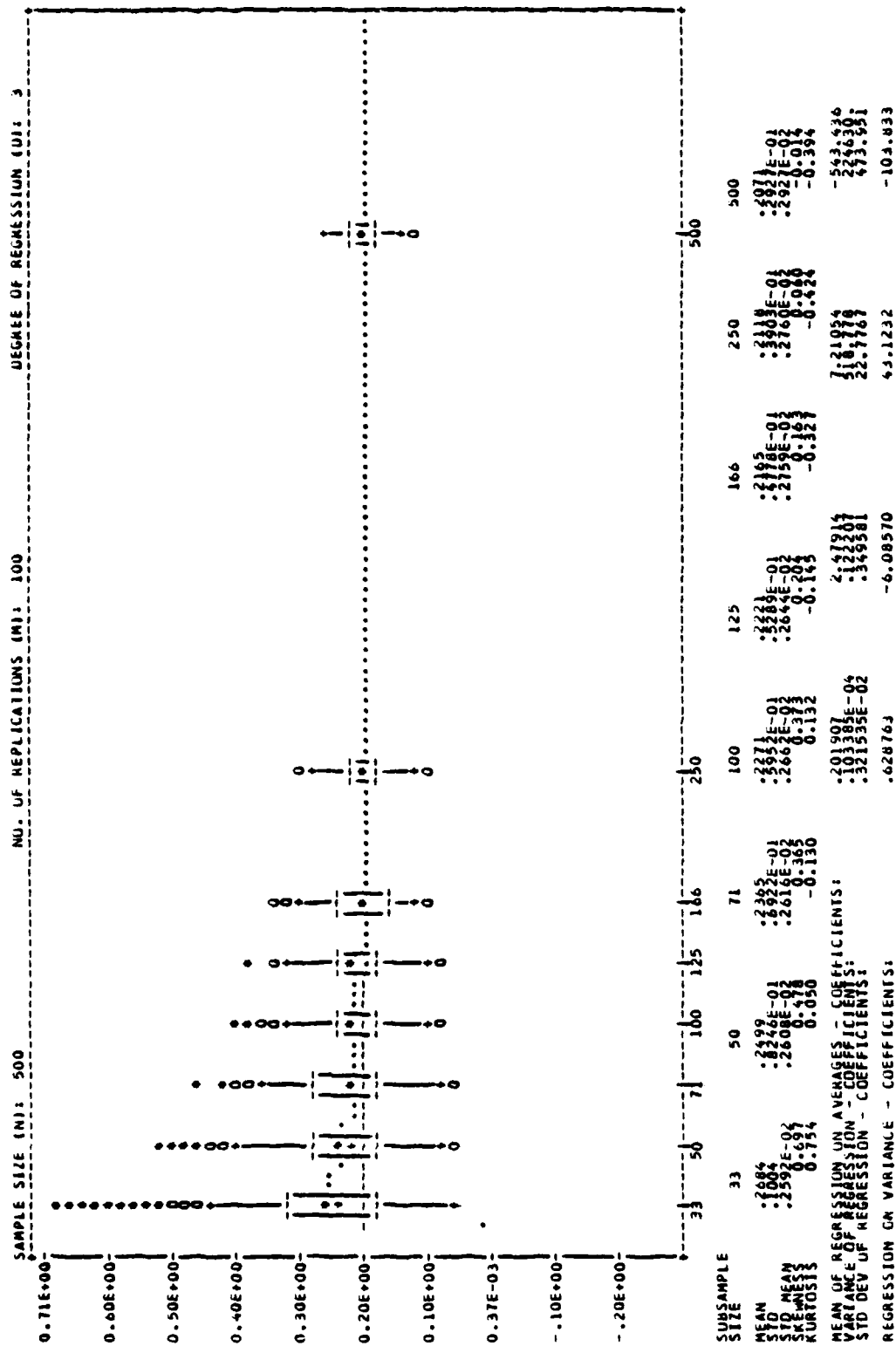
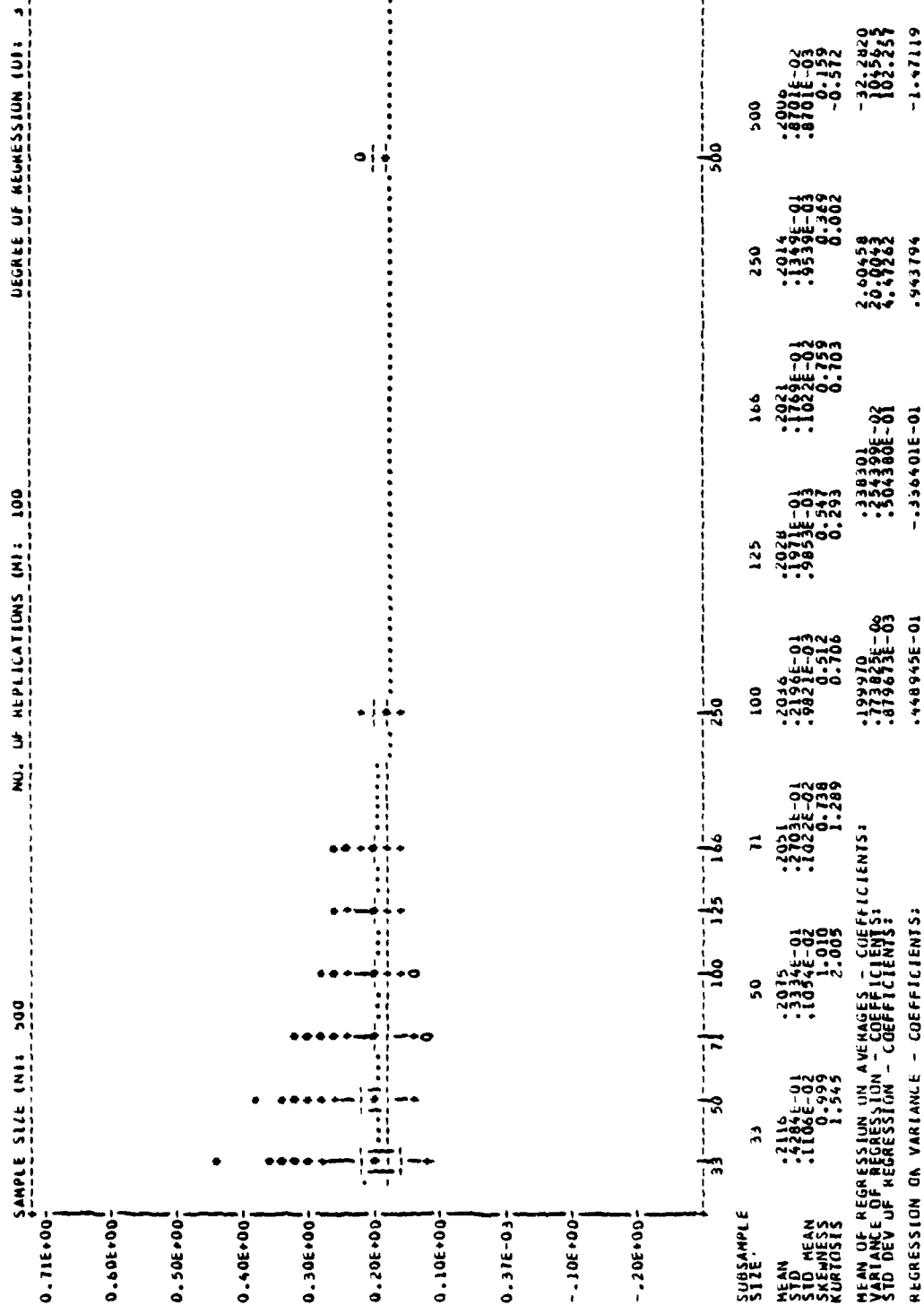


Figure 9a



VERTICAL SCALE: YMIN = -0.3815  
YMAX = 0.7050

ESTIMATOR: MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 9b

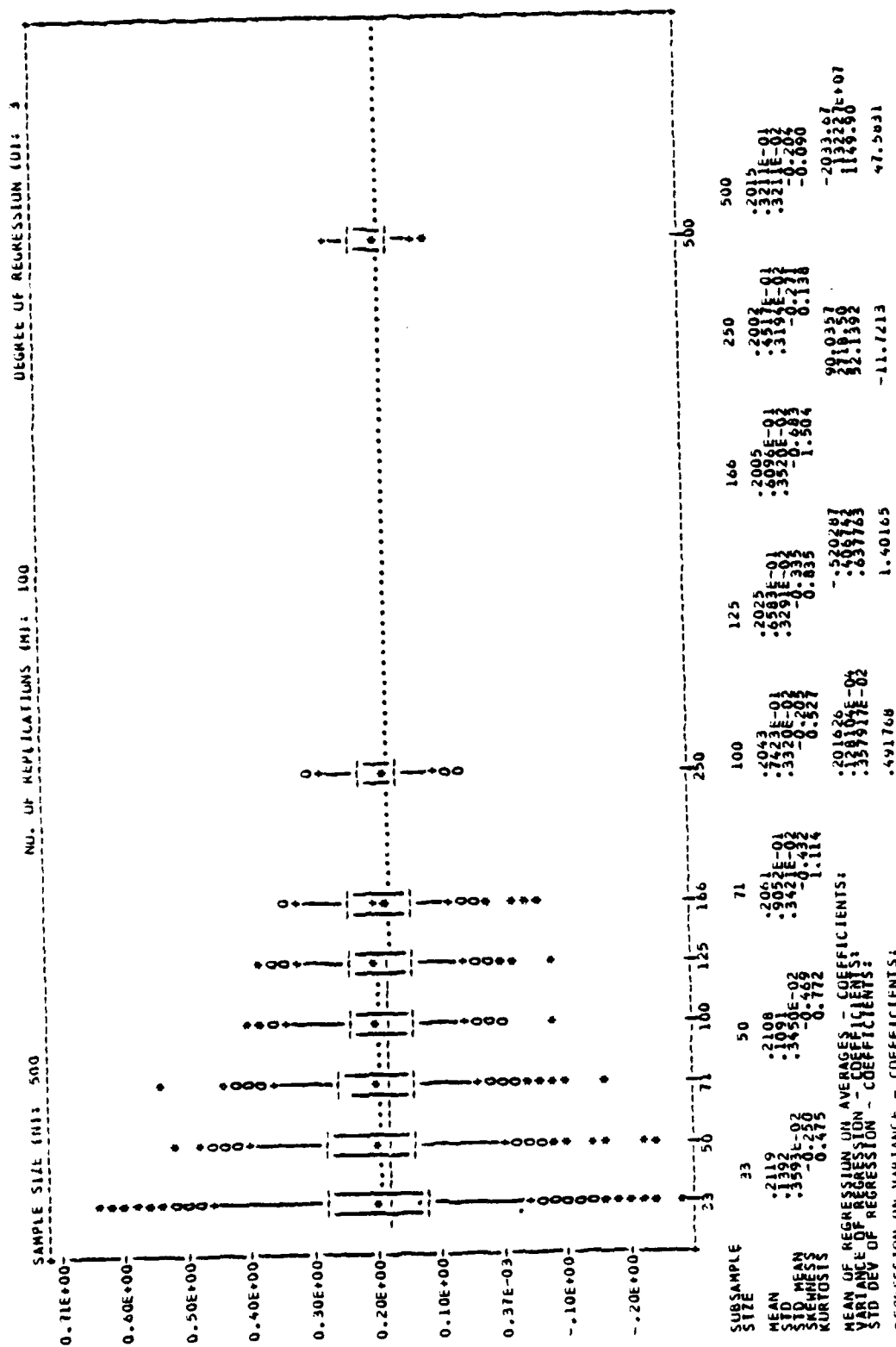
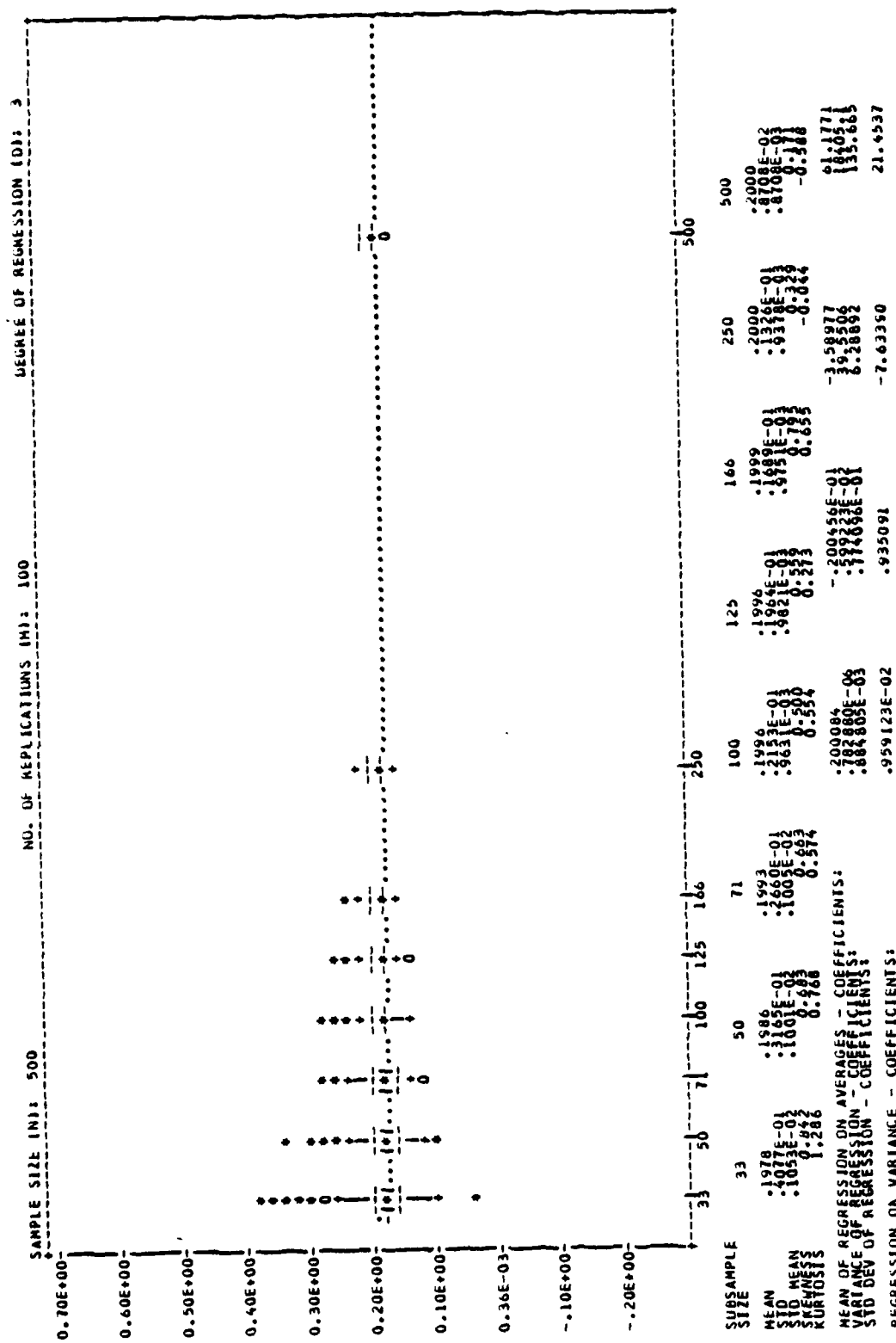
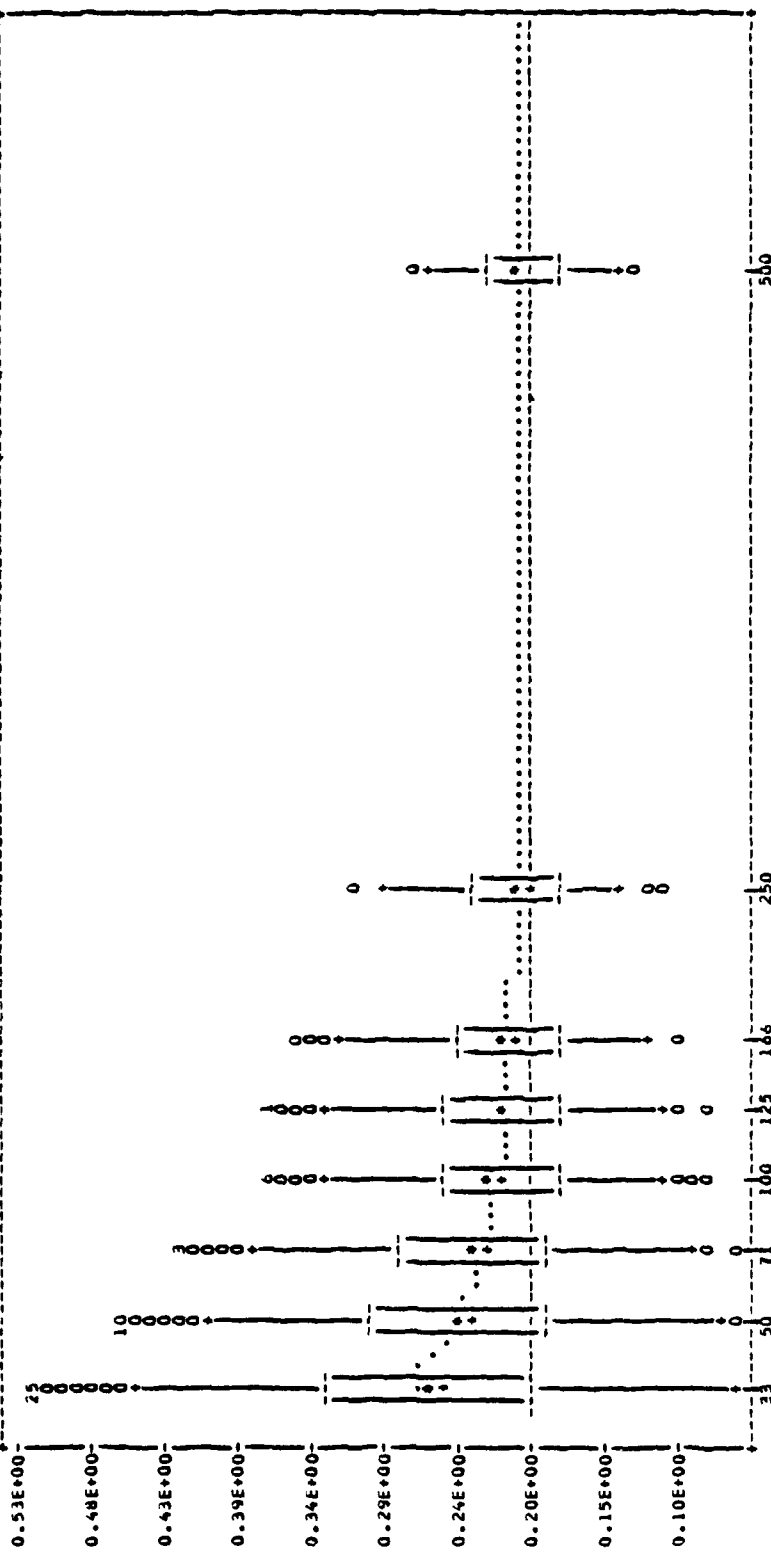


Figure 9c





SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	.2884	.2499	.2365	.2271	.2224	.2195	.2189	.2071
STD MEAN	.1004	.0246E-01	.0922E-01	.0934E-01	.0809E-01	.0770E-01	.0706E-01	.071E-01
STD MEAN	.2592E-02	.2808E-02	.2610E-02	.2668E-02	.2644E-02	.2759E-02	.2780E-02	.2921E-02
KURTOSIS	0.977	0.450	0.385	0.313	0.143	0.327	0.429	0.394
MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:				.201907	2.47914		7.21054	-543.436
VARIANCE OF REGRESSION - COEFFICIENTS:				.103385E-04	.122207		518.778	224.330
STD DEV OF REGRESSION - COEFFICIENTS:				.321555E-02	.349581		22.7767	473.951
REGRESSION ON VARIANCE - COEFFICIENTS:				.620763	-6.08570		43.1232	-103.633

VERTICAL SCALE: YMIN = 0.9529  
YMAX = 0.9286

ESTIMATOR: MOMENT ESTIMATOR (RECIPROCAL OF SQUARED COEFFICIENT OF VARIATION) OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.

Figure 10a (REDUCED GRAPHICS)



3. Case 2, Example 3: Large Subsample Sizes and Large Replications ( $K=0.2$ )

As a last example, graphs for large subsample sizes with a large number of replications ( $M=25$ ) are produced for each estimator. The box plots of these graphs (Figures 11a through 11d) look similar to their counterparts for small sample sizes and large replications (Figures 9a through 9d). This is due to the difference in the vertical scale and does not indicate a lack of variance reduction by larger sample sizes. The range of the vertical scale of Figures 11a through 11d is .495 and is smaller than that of Figures 9a through 9d, which is .9865, by a factor of two. This is a simple but important illustration of the caution that must be exercised when comparing graphics that are not of the same scale.

Other than the reduced variance caused by larger sample sizes, the same conclusions that were reached in Example 2 of this case concerning the variance and the bias of the estimators is shown to hold here. The MLE estimator shows considerably less variance and bias than the moment estimator. The jackknifed moment estimator results in more normally distributed estimates and less bias. The jackknifed MLE estimator shows some improvement toward normality of its estimates and its variance has changed little.

As in Case 1 for  $K=5.0$ , the larger the subsample size the more normal are the population of estimates. This

increased normality can be seen graphically by comparing the graphs of Figures 11a through 11d with those of Figures 9a through 9d. The lack of outliers on the high side of the box plots for Figures 9a to 9d is an indication of convergence normality.

Table 7 summarizes the results of the comparisons of the four different estimators for  $K=5.0$  and  $K=0.2$ . A comparison of the standard deviations shows the moment estimator to be consistently higher than the MLE estimator. These standard deviations, however, show little change when the primary estimators are jackknifed, indicating little or no change in variance by jackknifing. The bias values of the moment estimator are consistently higher and the root mean square error (r.m.s.e.) is considerably larger for  $K=0.2$  in comparison to the MLE estimator, indicating how poor the moment estimator is for evaluating shape parameters of gamma distribution that have small values.

These conclusions were easily drawn from the graphs and the statistical output provided by the RAGE program and did not require extensive programming and analysis by the experimenter. This kind of expedient examination of estimators not only saves the experimenter time but reduces significantly the computer time required for analysis. This facilitation of analysis is the primary purpose of the RAGE methodology.

Table 7: Bias and variance comparisons for varying subsample sizes and shape parameters

SHAPE PRMTR	SAMPLE & SUBSAMPLE SIZE	M	ESTIMATOR	ESTIMATE	STD.DEV.	BIAS	R.M.S.E.
0.2	500	100	MOMENT	.2071	.002927	-.0071	.03012
			MLE	.2006	.0008701	-.006	.01057
			JN MOM	.2015	.003211	-.0015	.03215
			JN MLE	.2000	.008707	.00	.0087
0.2	2000	25	MOMENT	.2021	.003584	-.0021	.01804
			MLE	.2001	.0008417	.0001	.00421
			JN MOM	.2005	.003644	-.0005	.01823
			JN MLE	.1997	.0008966	.0003	.00494
5.0	500	100	MOMENT	5.012	.03417	-.012	.3419
			MLE	5.095	.02730	-.095	.2891
			JN MOM	4.981	.0340	.019	.3405
			JN MLE	5.057	.02759	-.057	.2817
5.0	2000	25	MOMENT	4.994	.03184	.006	.1593
			MLE	5.049	.01821	-.049	.1033
			JN MOM	4.988	.03194	.012	.16015
			JN MLE	5.014	.02180	-.014	.1099

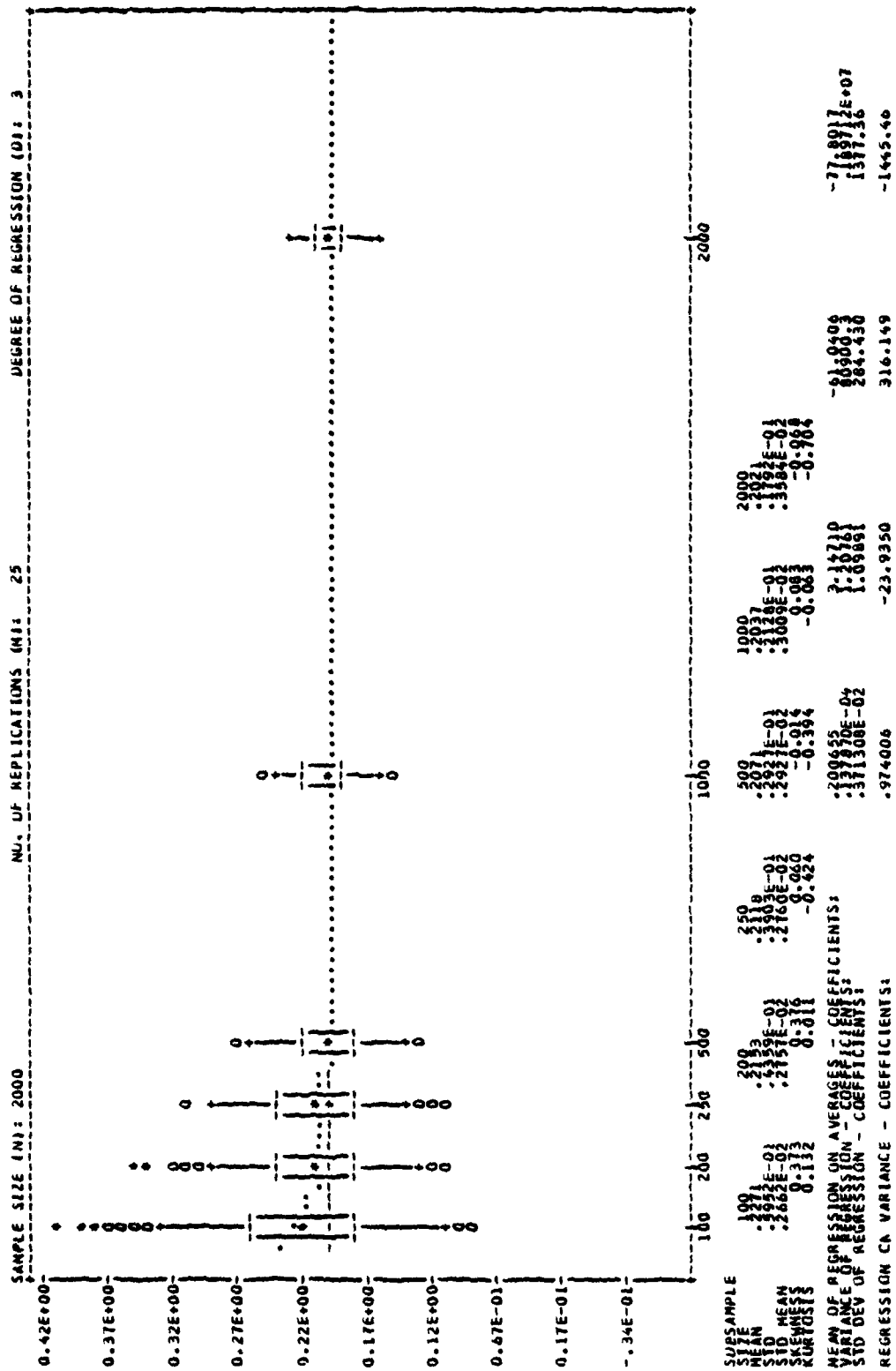
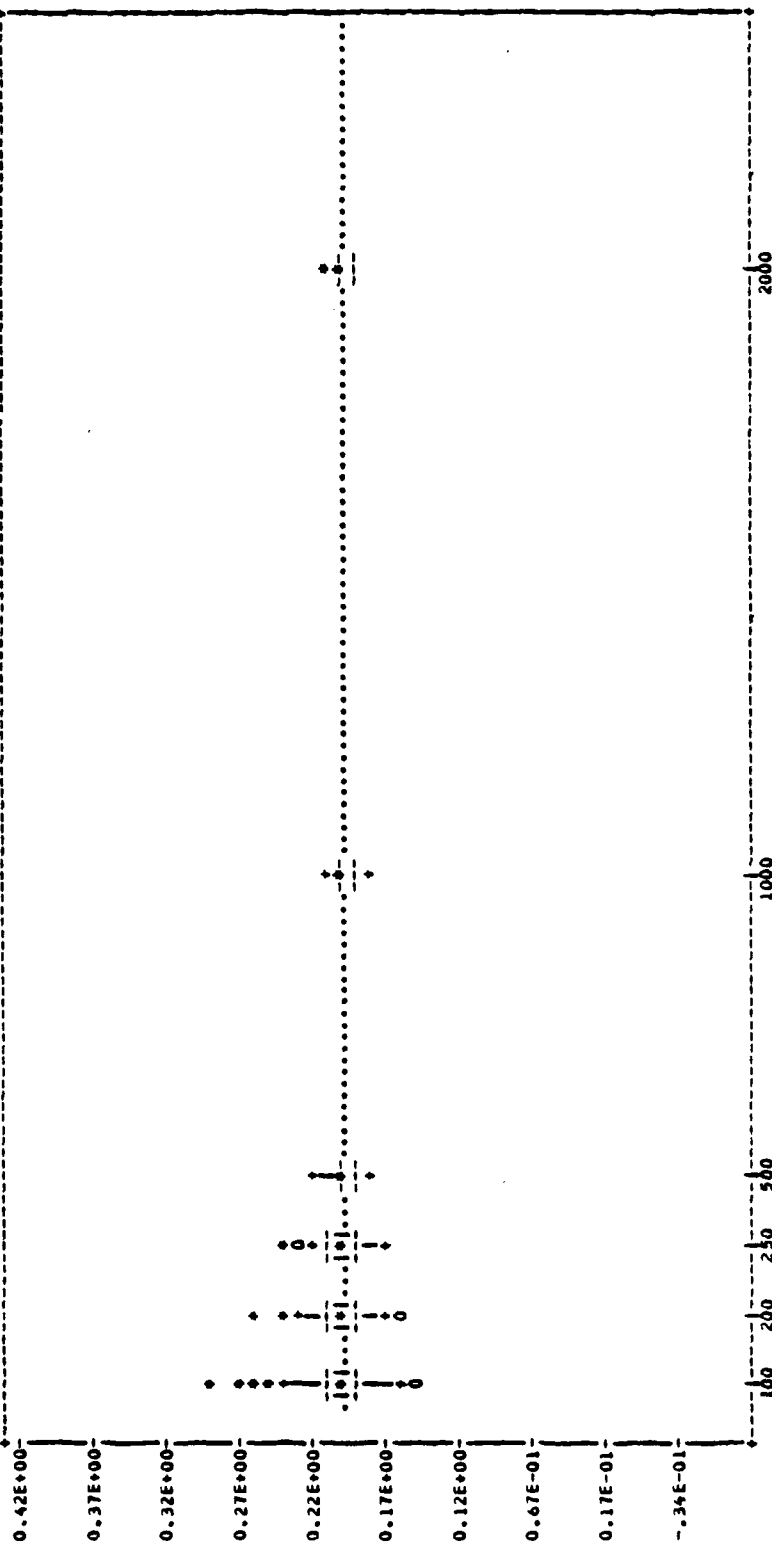


Figure 11a

SAMPLE SIZE (N): 2000

NU. OF REPLICATIONS (M): 25

DEGREE OF REGRESSION (D): 3



SUBSAMPLE SIZE	100	200	500	1000	2000
MEAN	.2036	.2017	.2009	.2004	.2001
STD MEAN	.2196E-01	.1307E-01	.8701E-03	.6815E-03	.5201E-03
STD DEV	.9821E-01	.9529E-01	.8701E-03	.6815E-03	.5201E-03
KURTOSIS	0.512	0.382	0.372	0.372	0.372
MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:			.19897	.41068	.110932
VARIANCE OF REGRESSION - COEFFICIENTS:			.68344E-04	.134374	.335370
STD DEV OF REGRESSION - COEFFICIENTS:			.82780E-03	.366443	16.3382
REGRESSION ON VARIANCE - COEFFICIENTS:			.530299E-01	-.860443	16.3382
VERTICAL SCALE: YMIN =					-58.7863
YMAX =					24837.6
ESTIMATOR: MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION.					127.589
					-89.6273

Figure 11b

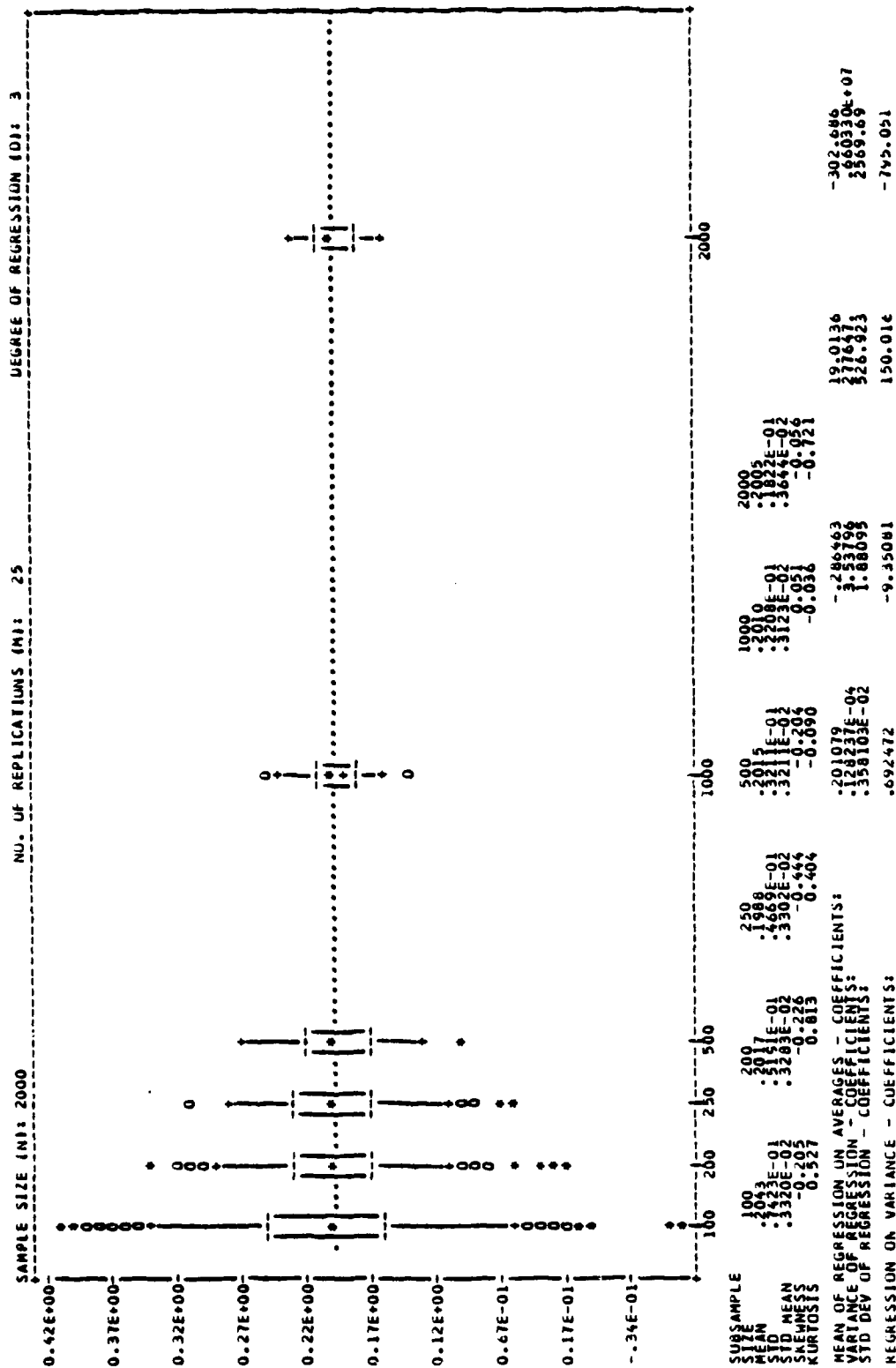


Figure 11c

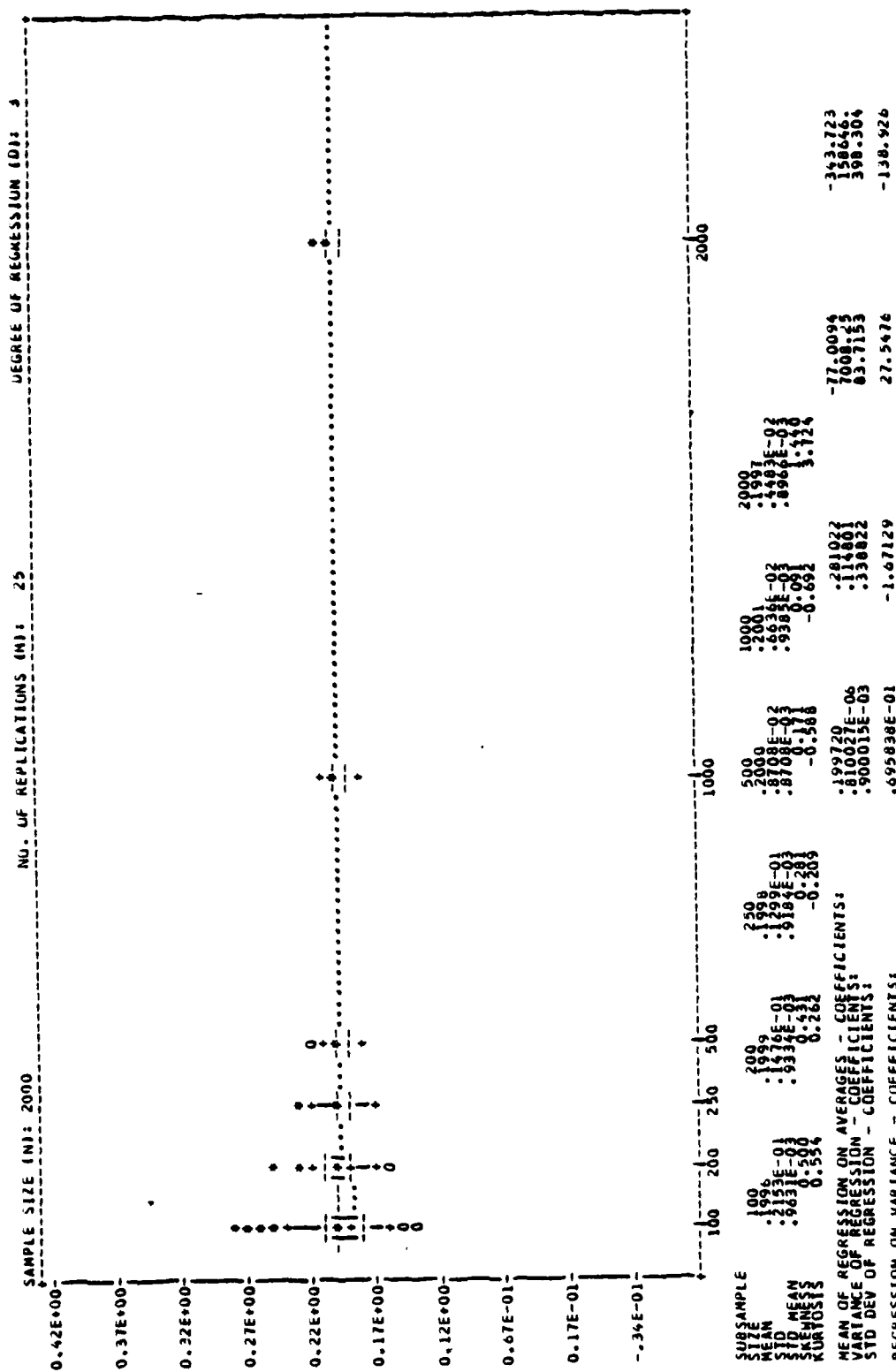


Figure 11d

# RAGE PROGRAM LISTING

\*\*\*\*\* RAGE SUBROUTINE \*\*\*\*\*

PURPOSE TO GENERATE REGRESSION ADJUSTED ESTIMATES AND BOX PLOTS  
OF ESTIMATES OF AN INPUT RAW DATA SERIES X CONTAINING M  
(REPLICATIONS) OF N VALUES EACH. UP TO 3 ESTIMATING  
FUNCTIONS CAN BE USED. THE GRAPHS CAN ALL BE OF THE SAME  
SCALE OR SCALED INDIVIDUALLY.

## DESCRIPTION OF PARAMETERS

X REAL\*4 ARRAY CONTAINING DATA.  
A MAXIMUM OF 50,000 DATA ELEMENTS CAN BE STORED IN X.

N NUMBER OF DATA ELEMENTS PER SECTION (N IS SAMPLE SIZE).  
N CANNOT EXCEED 50,000 AND M\*N MUST NOT EXCEED 50,000.

M NUMBER OF SECTIONS (REPLICATIONS).  
M CANNOT EXCEED 100 AND M\*N MUST NOT EXCEED 50,000.

NE INTEGER ARRAY OF SIZE 8 CONTAINING SUBSAMPLE SIZES FOR N.  
THE VALUES OF NE MUST BE FROM SMALLEST TO LARGEST.  
NO ELEMENT OF THE ARRAY NE CAN BE GREATER THAN N.  
M\*(N/NE(1)) MUST NOT EXCEED 12,500.

L NUMBER OF SUBSAMPLE SIZES FROM NE(8) THAT WILL BE USED TO  
SECTION N.

D IT IS ALSO THE NUMBER OF BOXPLOTS THAT WILL BE PRODUCED.  
DEGREE OF REGRESSION FOR MEAN AND VARIANCE REGRESSIONS.  
D WILL BE REDUCED BY RAGE IF THE SAMPLE IS NOT LARGE  
ENOUGH. D MUST BE 1, 2 OR 3.

\*\*\* SCALING \*\*\*  
SCALING IS ACCOMPLISHED BY TAKING THE SMALLEST SUBSAMPLE  
SIZE, NE(1), AND COMPUTING THE M\*(N/NE(1)) ESTIMATES.  
THE RANGE OF THESE VALUES IS THEN THE VERTICAL SCALE.  
IT IS ASSUMED THAT ESTIMATES FROM NE(2), NE(3), ETC  
SUBSAMPLE SIZES ARE WITHIN THE RANGE OF THOSE PRODUCED

BY NE(1) BECAUSE OF THE LARGER SAMPLE SIZES.  
 THE SEI PARAMETER ALLOWS THE USER TO SCALE THE GRAPHS  
 OF EACH ESTIMATOR INDIVIDUALLY OR TO SCALE THEM ALL TO  
 THE SAME SCALE. SCALING ALL TO THE SAME SCALE IS  
 ACCOMPLISHED BY TAKING THE MINIMUM AND MAXIMUM ESTIMATE  
 FROM ALL THE ESTIMATORS USING NE(1) SUBSAMPLE SIZE.  
 THE RG PARAMETER ALLOWS THE USER TO REDUCE THE VER-  
 TICAL SCALE TO: THE UPPER QUARTILE DISTANCE + 1.5 TIMES  
 INTERQUARTILE DISTANCE AS THE MAX VALUE AND THE LOWER  
 QUARTILE - 1.5 TIMES THE INTERQUARTILE DISTANCE AS THE  
 MIN VALUE. THE INTERQUARTILE DISTANCE IS COMPUTED FROM  
 THE SAMPLE OF ESTIMATES FROM THE NE(1) SUBSAMPLE SIZE.  
 IF THERE ARE NO ESTIMATES OUTSIDE THESE MIN AND MAX  
 VALUES THEN THE SCALE IS TO THE FIRST VALUE WITHIN.  
 IF THERE ARE ESTIMATES OUTSIDE THESE LIMITS THEN THEY  
 ARE COUNTED AND THE NUMBER PRINTED AT THE ENDS OF THE  
 BOX PLOTS.  
 THE SVS PARAMETER ALLOWS THE USER TO SET THE VERTICAL  
 SCALE. WHEN THE VERTICAL SCALE IS SET THE SEI PARAMETER  
 IS IGNORED AND THE VERTICAL SCALE BECOMES YMIN AND YMAX.

RG RG=0 DO NOT REDUCE THE VERTICAL SCALE OF THE GRAPHS.  
 RG=1 REDUCE GRAPHICS VERTICAL SCALE TO UPPER (LOWER)  
 QUARTILE + (-) 1.5 TIMES INTERQUARTILE DISTANCE.

SEI SEI=0 DO NOT SCALE ESTIMATORS' GRAPHS INDIVIDUALLY.  
 SEI=1 SCALE ESTIMATORS' GRAPHS INDIVIDUALLY.

SVS SVS=0 DO NOT SET THE VERTICAL SCALE.  
 SVS=1 SET VERTICAL SCALE TO YMIN AND YMAX.

YMIN MINIMUM VALUE OF VERTICAL SCALE.

YMAX MAXIMUM VALUE OF VERTICAL SCALE.

NEST NUMBER OF ESTIMATORS THAT WILL BE USED TO CALCULATE  
 STATISTICAL PARAMETER FROM X DATA.  
 NEST MUST BE 1, 2 OR 3.

EST1 NAMES OF THE ESTIMATOR FUNCTIONS THAT WILL BE USED TO  
 EST2 CALCULATE THE STATISTICAL PARAMETER.  
 EST3 CALL SEQUENCE ON EACH FUNCTION IS: CALL FNAME(X,N) WHERE  
 X IS THE DATA ARRAY AND N IS THE NUMBER OF DATA POINTS.  
 THEY MUST BE DECLARED IN THE CALLING PROGRAM (RAGE) IN  
 THE ORDER THEY ARE USED. DUMMY VARIABLES MUST BE INSERTED  
 WHEN THERE ARE LESS THAN 3 ESTIMATORS.

CC

TITLES ASSOCIATED WITH EACH ESTIMATOR (EST1,2,3). A MAX  
OF 120 CHARACTERS CAN BE USED TO DESCRIBE EACH ESTIMATOR.  
EACH TITLE MUST BE DECLARED AS REAL\*8(15) ARRAYS UNLESS  
PASSED AS AN ARGUMENT OF THE CALLING PROGRAM RAGE.  
WHEN PASSING THE TITLE AS AN ARGUMENT THERE MUST BE A  
MINIMUM OF 120 CHARACTERS BETWEEN APOSTROPHES.

TTL1  
TTL2  
TTL3

SUBROUTINE RAGE(X,N,M,NE,L,D,RG,SEI,SVS,YMIN,YMAX,NEST,EST1,

+TTL1,EST2,TTL2,EST3,TTL3)  
REAL X(50000),DLH(4),Y(12500)  
REAL\*8 TTL1(15),TTL2(15),TTL3(15)  
INTEGER NE(8),RG,SEI,SVS,SH  
INTEGER D,L,NEST,TEST

SM=NE(1)  
MN = H\*N

LT=L-1  
IF (LT.EQ.0) GO TO 13

DO 10 I=1,LT

I1=I+1

IF (NE(I1).GT.NE(I1)) WRITE(6,110)

10 CONTINUE

13 TEST=0

IF (NEST.EQ.1.OR.NEST.EQ.2.OR.NEST.EQ.3) GO TO 1

WRITE(6,106)

TEST=1

IF (MN.LE.50000) GO TO 2

WRITE(6,105)

TEST=1

IF (M.GE.1.AND.M.LE.100) GO TO 3

WRITE(6,104)

TEST=1

IF (L.GE.1.AND.L.LE.8) GO TO 4

WRITE(6,103)

TEST=1

IF (D.LE.3) GO TO 5

WRITE(6,108)

TEST=1

K=N/NE(L)

IF (K.GE.1) GO TO 6

WRITE(6,107)

TEST=1

K=M\*(N/NE(1))

IF (K.LE.12500) GO TO 7

WRITE(6,109)

TEST=1

IF (TEST.NE.0) GO TO 80

AD-A118 214

NAVAL POSTGRADUATE SCHOOL MONTEREY CA

F/G 12/1

A GRAPHICAL TEST BED FOR ANALYZING AND REPORTING THE RESULTS OF--ETC(U)

MAR 82 D G LINNEBUR

UNCLASSIFIED

NL

2 OF 2

AD-A  
118214



END  
DATE  
FILMED  
09-82  
DTIC

```

C      ULH(2)=YMIN
C      ULH(4)=YMAX
C      DETERMINE HOW EACH GRAPH IS TO BE SCALED.
C      IF (SVS.EQ.1) GO TO 50
C      IF (SEI.EQ.1) GO TO 75
C      *****
C      *GRAPH ALL ESTIMATORS TO THE SAME SCALE OF ESTIMATOR W/WIDEST PTS:
C      *****
C      PIND VERTICAL SCALE FOR 1ST ESTIMATOR.
C      CALL SECEST(X,N,M,SH,EST1,Y,KP)
C      IF (RG.EQ.1) CALL DELETO(Y,KP,YMAX,YMIN)
C      IF (RG.NE.1) CALL MAXMIN(Y,KP,YMAX,YMIN)
C      ULH(2)=YMIN
C      ULH(4)=YMAX
C      PIND VERTICAL SCALE FOR 2ND ESTIMATOR.  KEEP WIDEST PAIR****
C      IF (NEST.LT.2) GO TO 50
C      CALL SECEST(X,N,M,SH,EST2,Y,KP)
C      IF (RG.EQ.1) CALL DELETO(Y,KP,YMAX,YMIN)
C      IF (RG.NE.1) CALL MAXMIN(Y,KP,YMAX,YMIN)
C      IF (YMIN.LT.ULH(2)) ULH(2)=YMIN
C      IF (YMAX.GT.ULH(4)) ULH(4)=YMAX
C      PIND VERTICAL SCALE FOR 3RD ESTIMATOR.  KEEP WIDEST PAIR****
C      IF (NEST.LT.3) GO TO 50
C      CALL SECEST(X,N,M,SH,EST3,Y,KP)
C      IF (RG.EQ.1) CALL DELETO(Y,KP,YMAX,YMIN)
C      IF (RG.NE.1) CALL MAXMIN(Y,KP,YMAX,YMIN)
C      IF (YMIN.LT.ULH(2)) ULH(2)=YMIN
C      IF (YMAX.GT.ULH(4)) ULH(4)=YMAX
C      *****
C      PROCESS BOXPLOTS USING FIXED VERTICAL SCALE IN VECTOR ULH
C      ONE CALL FOR EACH ESTIMATOR USED.
C      CALL PRST(X,N,M,EST1,NE,L,RG,D,ULH,Y)
C      WRITE(6,101) ULH(2),ULH(4)
C      WRITE(6,102) TTI
C      IF (NEST.LT.2) GO TO 80
C      CALL PRST(X,N,M,EST2,NE,L,RG,D,ULH,Y)
C      WRITE(6,101) ULH(2),ULH(4)
C      WRITE(6,102) TTI
C      IF (NEST.LT.3) GO TO 80
C      CALL PRST(X,N,M,EST3,NE,L,RG,D,ULH,Y)
C      WRITE(6,101) ULH(2),ULH(4)
C      WRITE(6,102) TTI
C      GO TO 80

```

50

```

C C C
* GRAPH EACH ESTIMATOR SCALED TO ITS WIDEST POINTS.
*
* FIND VERTICAL SCALE FOR 1ST ESTIMATOR AND GRAPH.
75 CALL SECEST(X,N,M,NE(1),EST1,Y,KP)
   IF(RG.EQ.1) CALL DELETO(Y,KP,YMAX,YMIN)
   ULM(2)=YMIN
   ULM(4)=YMAX
   CALL PRST(X,N,M,EST1,NE,L,RG,D,ULH,Y)
   WRITE(6,101) TTL
   IF(NEST.LT.2) GO TO 80

C C
* FIND VERTICAL SCALE FOR 2ND ESTIMATOR AND GRAPH.
   CALL SECEST(X,N,M,NE(1),EST2,Y,KP)
   IF(RG.EQ.1) CALL DELETO(Y,KP,YMAX,YMIN)
   ULM(2)=YMIN
   ULM(4)=YMAX
   CALL PRST(X,N,M,EST2,NE,L,RG,D,ULH,Y)
   WRITE(6,101) TTL
   IF(NEST.LT.3) GO TO 80

C C
* FIND VERTICAL SCALE FOR 3RD ESTIMATOR AND GRAPH.
   IF(SVS.EQ.1) GO TO 78
   CALL SECEST(X,N,M,NE(1),EST3,Y,KP)
   IF(RG.EQ.1) CALL DELETO(Y,KP,YMAX,YMIN)
   ULM(2)=YMIN
   ULM(4)=YMAX
   CALL PRST(X,N,M,EST3,NE,L,RG,D,ULH,Y)
   WRITE(6,101) TTL

C
80 RETURN
102 FORMAT(1X,'ESTIMATOR: ',15A8)
103 FORMAT(1X,'VERTICAL SCALE: YMIN =',F10.4,'/18X,YMAX =',F10.4)
104 FORMAT(1X,'*** ERROR...L MUST BE AN INTEGER BETWEEN 1 AND 8.***')
105 FORMAT(1X,'*** ERROR...M MUST BE AN INTEGER BETWEEN 1 AND 100.***')
106 FORMAT(1X,'*** ERROR...N EXCEEDS 50,000.***')
107 FORMAT(1X,'*** ERROR...NEST MUST BE 3 OR LESS.***')
108 FORMAT(1X,'*** ERROR...N/NE(L) MUST BE 1 OR GREATER TO COMPUTE.')
109 FORMAT(1X,'*** ERROR...D INCREASE N OR DECREASE NE(L) TO 3.***')
110 FORMAT(1X,'*** ERROR...M*(N/NE(1)) MUST NOT EXCEED 12,500.***')
      + STATISTICS.
      * NE ARRAY ELEMENTS ARE NOT IN ORDER OF ./

```

```

+! INCREASING SIZE. IF NE(1) IS NOT SMALLEST ELEMENT, SCALING'/
+! MAY CAUSE POINTS TO FALL OUTSIDE RANGE OF SCALE.!!
END
***** PRST SUBROUTINE *****
SUBROUTINE PRST(X,N,M,EST,NE,L,RG,UD,ULH,Y)
REGRESSION ADJUSTED ESTIMATE
CALCULATES ESTIMATES FROM USER DATA USING "EST" FUNCTION
PLOTS BASIC OR RETRENCHED GRAPH ON LINE PRINTER

EST = NAME OF USER WRITTEN ESTIMATING FUNCTION.
USAGE: FUNCTION NAME(X,N)
WHERE X IS A VECTOR WITH N ENTRIES

M = NO. OF REPLICATIONS (MUST BE <= 100)
N = NUMBER OF VALUES IN EACH REPLICATION (M*N MUST BE <= 50000)
X = USERS VECTOR WITH M CONSECUTIVE BATCHES OF N VALUES EACH
L = NO. OF SECTION SIZES (MUST BE BETWEEN 1 AND 8)
NE = ARRAY WITH THE L SUBSAMPLE SIZES (MUST BE IN ASCENDING ORDER)
UD = DEGREE OF THE REGRESSION (MUST BE <= 3 & <= L-1)
ULH = XMIN, YMIN, XMAX, YMAX IN USER UNITS
      ONLY ULH(2) AND ULH(4) NEED TO BE PASSED. OTHERS CALC. HERE

DIMENSION NE(L)
INTEGER NB(8), LDCX(8), D1, IWIDTH, UD, D, LT, DT, RG, RNEK
INTEGER*2 PLOT(122,50), CBAR, BLK, DASH, CSTR, NUH(10), DOT
REAL*4 X(50000), ULH(4), DLH(4), Y(12500)
REAL*4 RH(8,100), STAT(8,6), VT(8)
REAL*8 SUM2, SUM3, SUM4, LABEL(5)
REAL*4 RA(8,4), RV(8,4), B(4,100), V(4), BA(4), BV(4), BS(4), BT(8), BT(8)
DATA DLH/1.1, 1.1, 122.50, /
DATA BLK/1.1, 1.1, DASH/1.1, 1.1, /
DATA LABEL/'MEAN', 'STD', 'STD MEAN', 'SKEWNESS', 'KURTOSIS' /
D=MINO(3,UD,L-1)
IX1=8
IX2=4
MN=M*N
D1=D+1
IWIDTH= IFIX(DLH(3))
BUILD REGRESSION MATRICES FOR AVERAGES AND VARIANCES
DO 84 K=1,L
DO 86 J=1,D1
T=NE(K)/FLOAT(N)
RA(K,J)=1./T**J-1
RV(K,J)=1./T**J(J+1)/2.)
86 CONTINUE
84 CONTINUE

```

```

C      CLEAR PLOT ARRAY
DO 3 J=1,50
DO 4 I=1,122
  PLOT(I,J)=BLK
  CONTINUE
3    SET HORIZONTAL XMIN, XMAX
  ULH(1)=1.7*NE(1)
  ULH(3)=1.2*NE(L)
  SET SCALE
  CALL SCALE(ULH,DLH)
  COMPUTE LOCATION OF BOXPLOTS ALONG X-AXIS
  LAST=-1
DO 5 K=1,L
  NB(K)=N/NE(K)
  LOCI(K)=(NE(K)-ULH(1))* (DLH(3)-DLH(1))/(ULH(3)-ULH(1)) + 1
  IF(LOCI(K).LT. LAST+6) LOCI(K)=LAST+6
  LAST=LOCI(K)
5  CONTINUE

C      DO 80 K=1,L
  NBK=NB(K)
  RMK=NE(K)
  SECTION 8 COMPUTE ESTIMATORS FOR SIZE K
  CALL SECEST(X,N,M,RMK,EST,Y,KP)
  AVERAGE ESTIMATES OF SIZE NE(K) FOR EACH OF M REPLICATIONS
  KP=0
DO 10 I=1,M
  RH(K,I)=0
DO 15 J=1,NBK
  KP=KP+1
  RH(K,I)=RH(K,I)+Y(KP)
  CONTINUE
15  CONTINUE
  RH(K,I)=RH(K,I)/FLOAT(NBK)
10  CALL BOXPRT(Y,KP,LOCI(K),PLOT,RG)
  IF(K.GT.8) GO TO 80
  COMPUTE MEAN AND MOMENT ESTIMATES
  XMEAN=0.
DO 180 IN1=1,KP
  XMEAN=XMEAN+Y(IN1)
180 CONTINUE
  XMEAN=XMEAN/FLOAT(KP)
  SUM2 = 0.0D0
  SUM3 = 0.0D0
  SUM4 = 0.0D0
DO 190 IP1=1,KP
  DEV = Y(IP1) - XMEAN

```

```

SUM2 = SUM2 + DEV * DEV
SUM3 = SUM3 + DEV ** 3
SUM4 = SUM4 + DEV ** 4
190 CONTINUE
C
C
CHECK TO INSURE SAMPLE SIZE IS LARGE ENOUGH FOR
EACH MOMENT COMPUTATION.
IF (KP.LT.2) GO TO 7
VAR = SUM2 / (KP - 1.0)
STDV = SQRT(VAR)
7 IF (KP.LT.3) GO TO 8
XSUM3 = SNGL(SUM3) * KP / ((KP-1.) * (KP-2.))
SKEN = XSUM3 / STDV ** 3
8 IF (KP.LT.4) GO TO 9
XSUM4 = SUM4 * ((KP-2.) * KP + 3.) / ((KP-1.) * (KP-2.) * (KP-3.))
XSUM4 = XSUM4 - (VAR * VAR) * 3. / (KP * VAR) - 3.
CKURT = XSUM4 / (VAR * VAR)
9 STAT(K,1) = XMEAN
STAT(K,2) = STDV
STAT(K,3) = STDV / SQRT(FLOAT(KP))
STAT(K,4) = SKEN
STAT(K,5) = CKURT
STAT(K,6) = VAR
80 CONTINUE
C
C
IF D1.LT.2 THEN NO REGRESSIONS OR PLOTTING CAN BE DONE
IP(D1.LT.2) GO TO 113
DO 92 K=1,H
DO 47 J=1,L
RT(J) = RA(J,K)
47 CONTINUE
CALL RRREG(RA,RT,BT,L,D1,IX1,IX2)
B(1,K) = BT(1)
DO 23 KT=2,D1
B(KT,K) = BT(KT) * N ** (KT-1)
23 CONTINUE
92 CONTINUE
C
C
AVERAGE REGRESSION COEFF. OVER M REPLICATIONS & CALC. VARIANCE
DO 94 I=1,D1
BA(I) = 0.
BV(I) = 0.
DO 95 J=1,M
BA(I) = BA(I) + B(I,J)
BV(I) = BV(I) + B(I,J) ** 2
95 CONTINUE

```



```

C C C PLOT *****
113 WRITE(6,102) N,M,D
    WRITE(6,161)
    WRITE(6,101)
    DO 90 J=1,50
      K=51-J
      IF(MOD(K,5).NE.0) GO TO 85
      YLABEL=(K-1)*(ULH(4)-ULH(2))/(DLH(4)-DLH(2)) + ULH(2)
      WRITE(6,103) YLABEL, (PLOT(I,K), I=1, INIDTH)
      GO TO 90
85 CONTINUE
90 WRITE(6,100) (PLOT(I,K), I=1, INIDTH)
    CONTINUE
    LABEL X-Axis BY REUSING PLOT MATRIX.
    DO 115 I=1,122
      PLOT(I,1) = DASH
      PLOT(I,2) = BLK
115 CONTINUE
    DO 130 J=1, L
      PLOT(LOCX(J), 1) = CBAR
      IX = NE(J)
      CALL NUMPRT(IX, 2, IK, PLOT)
130 CONTINUE
      WRITE(6,106) (PLOT(I,1), I=1, INIDTH)
      WRITE(6,104) (PLOT(I,2), I=1, INIDTH)
      L8=L
      IF(L8.GT.8) L8=8
      WRITE(6,156) (NE(I), I=1, L8)
      WRITE(6,146) LABEL(1), STAT(K,1), K=1, L8)
      CHECK TO INSURE SAMPLE SIZE IS LARGE ENOUGH TO COMPUTE
      STATISTICS BEFORE ATTEMPTING TO PRINT STATS.
      LT=L8
      DO 21 I=1, LT
        I=1, LT
        K1 = N/NE(L1)
        IF (K1.GE.2) GO TO 11
        L1 = L1 - 1
21 CONTINUE
11 GO TO 14
    WRITE(6,157) LABEL(2), (STAT(K,3), K=1, L1)
    WRITE(6,157) LABEL(3), (STAT(K,3), K=1, L1)
    LT = L1
    DO 22 I=1, LT
      K1 = N/NE(L1)

```

```

      IF (K1.GE.3) GO TO 12
      L1 = L1-1
22  CONTINUE
      GO TO 14
12  WRITE(6,158) LABEL(4), (STAT(K,4),K=1,L1)
      LT = L1
      DO 33 I=1,LT
        K1 = N/NE(L1)
        IF (K1.GE.4) GO TO 13
        L1 = L1-1
33  CONTINUE
      GO TO 14
13  WRITE(6,158) LABEL(5), (STAT(K,5),K=1,L1)
      IF(D1.LT.2) GO TO 44
      WRITE(6,151) (BA(I),I=1,D1)
14  IF(M.LT.2) GO TO 44
      WRITE(6,152) (BV(I),I=1,D1)
      WRITE(6,153) (BS(I),I=1,D1)
44  IF(DT.LE.1) GO TO 99
      IF(DT.LT.1) (V(I),I=1,DT)
99  WRITE(6,159)
100 FORMAT(9X,1,1,122A1,1,1)
101 FORMAT(9X,1,1,122A1,1,1)
102 FORMAT(11,1,1,122A1,1,1)
103 FORMAT(11,1,1,122A1,1,1)
104 FORMAT(10X,1,1,122A1,1,1)
106 FORMAT(9X,1,1,122A1,1,1)
107 FORMAT(10X,1,1,122A1,1,1)
151 FORMAT(1,1,1,122A1,1,1)
152 FORMAT(1,1,1,122A1,1,1)
153 FORMAT(1,1,1,122A1,1,1)
156 FORMAT(1,1,1,122A1,1,1)
146 FORMAT(1,1,1,122A1,1,1)
157 FORMAT(1X,A8,8F14.4)
158 FORMAT(1X,A8,8F14.3)
159 FORMAT(1,1,1,122A1,1,1)
161 FORMAT(9X,1,1,122A1,1,1)
162 FORMAT(15,18X,1,1,122A1,1,1)
206 FORMAT(5F16.8)
      RETURN
      END
C*****
C SUBROUTINE BOXPRT(Y,NY,IX,PLOT,RG)
C PREPARES BOXPLOT FROM VECTOR Y (IN 2-D ARRAY PLOT)
C REAL Y(NY),ULH(4),DLH(4)
C INTEGER RG

```

```

C
INTEGER*2 PLOT(122,50),DASH,CBAR,CROSS,CSTR,CO,NUM(10)
LOGICAL*1 LFLAG
DATA DASH//'-'/,CBAR//'.'/,CSTR//'+',CROSS//'+',CO//'0'//
IF(NY.GE.9) GO TO 5
WHEN LESS THAN 9 POINTS JUST SHOW THE POINTS
DO 8 I=1,NY
J=(Y(I)-YMIN)*VSCALE+1.
IGNORE VALUE IF IT FALLS OUTSIDE WINDOW
IF(J.GT.50.OR. J.LT.1) GO TO 8
PLOT(IX,J)=CO
8 CONTINUE
SUM=0.
DO 88 I=1,NY
SUM=SUM+Y(I)
88 CONTINUE
SUM=SUM/FLOAT(NY)
MEAN=(SUM-YMIN)*VSCALE+1
PLOT(IX,MEAN)=CSTR
GO TO 99
5 CONTINUE
LFLAG=.FALSE.
P25 = PCTILE(Y,NY,.25)
P75 = PCTILE(Y,NY,.75)
P50 = PCTILE(Y,NY,.50)
IQ1=(P25-YMIN)*VSCALE+1.
IQ3=(P75-YMIN)*VSCALE+1.
XLOW=2*(P25-P75)
XHI=2*(P75-P25)
XHI=(XHI-YMIN)*VSCALE+1.
XLOW=(XLOW-YMIN)*VSCALE+1.
THI=(XHI-YMIN)*VSCALE+1.
CLOW=2.5*(P25-1.5*P75)
CHI=2.5*(P75-1.5*P25)
DRAW BOX
DO 20 I=IQ1,IQ3
PLOT(IX-1,I)=CBAR
PLOT(IX+1,I)=CBAR
20 CONTINUE
PLOT(IX-1,IQ1)=DASH
PLOT(IX+1,IQ1)=DASH
PLOT(IX-1,IQ3)=DASH
PLOT(IX+1,IQ3)=DASH
DETERMINE IF OUTLIERS ARE TO BE COUNTED AND THE NUMBER PRINTED.
IF (RG.EQ.1) GO TO 55
DO 30 I=1,NY
J=(Y(I)-YMIN)*VSCALE+1.
IGNORE VALUE IF IT FALLS OUTSIDE WINDOW
IF(J.GT.50 .OR. J.LT.1) GO TO 30

```

```

C      IF(Y(I).LT.CLOW) PLOT(IX,J)=CSTR
      IF(Y(I).LT.CLOW) GO TO 36
      IF(Y(I).GE.CLOW .AND. Y(I).LT.XLOW) PLOT(IX,J)=CO
      IF(LFLAG .OR. Y(I).LT.XLOW) GO TO 25
      THIS IS THE LOW-CROSS POINTER (1ST POINT GE XLOW)
      LFLAG=.TRUE.
      ILX=J
C      NEXT LINE ENDS UP WITH HI-CROSS POINTER (LAST POINT LE XHI)
      IF(Y(I).LE.XHI) IHX=J
      IF(Y(I).GT.XHI .AND. Y(I).LE.CHI) PLOT(IX,J)=CO
      IF(Y(I).GT.CHI) PLOT(IX,J)=CSTR
      GO TO 56
C      SCALE TO INTERQUARTILE + (-) INTERQUARTILE DISTANCE.
C
C      II=0
C      III=0
      DO 31 I=1,NY
      J=(Y(I)-YMIN)*VSCALE+1
      IF(Y(I).LT.CLOW) II=II+1
      IF(Y(I).GT.CHI) III=III+1
      IF(Y(I).GT.50 .OR. J.LT.1) GO TO 31
      IF(Y(I).GE.CLOW .AND. Y(I).LT.XLOW) PLOT(IX,J)=CO
      IF(LFLAG .OR. Y(I).LT.XLOW) GO TO 26
      THIS IS THE LOW-CROSS POINTER (1ST POINT GE XLOW)
      LFLAG=.TRUE.
      ILX=J
C      NEXT LINE ENDS UP WITH HI-CROSS POINTER (LAST POINT LE XHI)
      IF(Y(I).LE.XHI) IHX=J
      IF(Y(I).GT.XHI .AND. Y(I).LE.CHI) PLOT(IX,J)=CO
      GO TO 31
C      PRINT NUMBER OF OUTLIERS UNLESS 0.
      DO 22 K=1,2
      IK=I
      J=(CLOW-YMIN)*VSCALE + 1
      IF(J.LT.0) J=1
      IF(K.EQ.2) IK=III
      IF(K.EQ.2) J=(CHI-YMIN)*VSCALE + 1
      IF(J.GT.50) J=50
      IF(IK.EQ.0) GO TO 22
      CALL NUMERT(IX,J,IK,PLOT)
      GO TO 22
C      CONTINUE
C
C      FILL BARS ABOVE AND BELOW THE BOX
      DO 32 I=ILX,IQ1
      PLOT(IX,I)=CBAR

```

```

32 CONTINUE
DO 33 I=IQ3, IHX
  PLOT(IX, I)=CBAR
33 CONTINUE
  PLOT(IY, IQ3)=CROSS
  PLOT(IY, IQ1)=DASH
  PLOT(IY, IQ2)=CROSS
  PLOT(IY, IQ3)=DASH
  SUM=0.
DO 40 I=1, NY
  SUM=SUM+Y(I)
40 CONTINUE
  SUM=SUM/FLOAT(NY)*VSCALE+1
  MEAN=(SUM-YMIN)*VSCALE+1
  PLOT(IY, MEAN)=CSTR
99 CONTINUE
RETURN
C ***** SCALE *****
ENTRY SCALE(ULH, DLH)
COMPUTES X, Y SCALE AND LIMITS
XMIN=ULH(1)
XMAX=ULH(3)
YMIN=ULH(2)
YMAX=ULH(4)
HSCALE=(DLH(3)-DLH(1))/(ULH(3)-ULH(1))
VSCALE=(DLH(4)-DLH(2))/(ULH(4)-ULH(2))
RETURN
END
C ***** RREG SUBROUTINE *****
C
C SUBROUTINE RREG(X, Y, B, M, N, IX1, IX2)
C ROBUST REGRESSION ON Y=X*B
C X=M BY N MATRIX CONTAINED IN AN ARRAY OF DIM(IX1, IX2)
C Y=M-VECTOR CONTAINED IN AN ARRAY OF DIM(IX1)
C B=N-VECTOR CONTAINED IN AN ARRAY OF DIM(IX2)
C XY, X1=WORK ARRAYS OF DIM(IX2, IX2)
C WY=WORK ARRAY OF DIM(IX1, IX2)
C WX=WORK MATRIX OF DIM(IX1, IX2)
C XY=WORK ARRAY OF DIM(N*2 + 3*N) OR LARGER
C WK=WORK ARRAY OF DIM(N*2 + 3*N) OR LARGER
C
C REAL*4 B(4), XY(4), WY(8)
C REAL*4 B(8), X(4), WK(500), WX(8,4)
C SUPPRESS ERROR MESSAGES: UNDERFLOW AND IMSL
C CALL ERRSET(208, 300, -1, 1)
C CALL UERSET(0, LEVOLD)
C K=1

```

```

L=0
CALL VMULPM(X,Y,M,N,N,IX1,IX1,XY,IX2,IER)
CALL VMULPM(X,Y,M,N,K,IX1,IX1,XY,IX2,IER)
CALL LINV2F(XY,M,IX1,IX1,IER)
CALL VMULFF(XY,XY,M,N,K,IX2,IX2,B,IX2,IER)
RETURN
END

```

```

C***** SUBROUTINE NUMPRT (IX,J,IK,PLOT) *****
C

```

```

C NUMPRT PLOTS THE NUMBER IK IN THE 2-D ARRAY PLOT CENTERED ON
C THE PLOT(IX,J) POSITION.
C IX = COLUMN OF MATRIX PLOT WHERE NUMBER IS TO BE PRINTED.
C J = ROW OF MATRIX WHERE NUMBER IS TO BE PRINTED.
C IK = NUMBER TO BE PRINTED
C PLOT = 2-D ARRAY WHERE NUMBER IS TO BE PLOTTED.
C

```

```

INTEGER*2 NUM(10), PLOT(122,50)
DATA NUM/0,1,2,3,4,5,6,7,8,9,/
IF (IK.LT.10) GO TO 1
IF (IK.LT.100) GO TO 2
IF (IK.LT.1000) GO TO 3
IF (IK.LT.10000) GO TO 4
I1000 = IK/10000
PLOT(IX-2,J) = NUM(I10000+1)
I1000 = (IK-I10000*10000)/1000
PLOT(IX-1,J) = NUM(I1000+1)
I100 = (IK-I10000*1000-I1000*1000)/100
PLOT(IX,J) = NUM(I100+1)
I10 = (IK-I10000*1000-I1000*1000-I1000*100)/10
PLOT(IX+1,J) = NUM(I10+1)
I1 = (IK-I10000*1000-I1000*1000-I1000*100-I10*10)
PLOT(IX+2,J) = NUM(I1+1)
GO TO 22

```

```

4 I1000 = IK/1000
PLOT(IX-2,J) = NUM(I1000+1)
I100 = (IK-I1000*100)/100
PLOT(IX-1,J) = NUM(I100+1)
I10 = (IK-I1000*100-I100*100)/10
PLOT(IX,J) = NUM(I10+1)
I1 = (IK-I1000*100-I100*100-I10*10)
PLOT(IX+1,J) = NUM(I1+1)
GO TO 22

```

```

3 I100 = IK/100
PLOT(IX-1,J) = NUM(I100+1)
I10 = (IK-I100*100)/10
PLOT(IX,J) = NUM(I10+1)
I1 = (IK-I100*100-I10*10)

```

```

PLOT (IX+1,J) = NUM (I1+1)
GO TO 22
2 I10 = IK/10
PLOT (IX-1,J) = NUM (I10+1)
I1 = (IK-I10*10)
PLOT (IX,J) = NUM (I1+1)
GO TO 22
1 PLOT (IX,J) = NUM (IK+1)
22 RETURN
END
*****
SUBROUTINE SECEST (X,N,M,NEK,EST,KP)
COMPUTE ESTIMATES "EST" FOR SECTION LENGTH NEK
REAL X(50000),Y(12500)
NBK=N/NEK
KP=0
DO 10 I=1,M
  IP= (I-1)*N + 1
  DO 15 J=1,NBK
    KP=KP+1
    Y(KP)=EST(X(IP),NEK)
    IP=IP+NEK
  15 CONTINUE
  10 CONTINUE
  RETURN
END
*****
SUBROUTINE MAXMIN (Y,N,YMAX,YMIN)
RETURNS MAX AND MIN VALUES OF VECTOR Y OF LENGTH N
REAL Y(N)
YMAX=Y(1)
YMIN=Y(1)
DO 605 J=1,N
  IF (Y(J).LT. YMIN) YMIN=Y(J)
  IF (Y(J).GT. YMAX) YMAX=Y(J)
  605 CONTINUE
RETURN
END
*****
FUNCTION PCTILE (Y,N,P)
COMPUTES P PERCENTILE OF N VALUES IN Y
REAL Y(N)
R=P*FLOAT(N+1)
CALL PXSORT (Y,1,N)
I=HAXO (INT(R),1)

```

```

I=MINO(I,N)
J=MINO(INT(R+1.),N)
R=R-INT(R)
PCTILE=Y(1)+R*(Y(J)-Y(I))
RETURN
END
C***** DELETO SUBROUTINE *****
C SUBROUTINE DELETO(Y,KP,YMAX,YMIN)
C SUBROUTINE SCALES THE GRAPH TO UPPER (LOWER) QUARTILE + (-)
C 1.5 TIMES INTERQUARTILE DISTANCE OR TO FIRST POINT WITHIN
C THESE LIMITS IF NO POINTS EXIST OUTSIDE.
C REAL Y(KP),Z(12500)
C DO 23 I=1,KP
C   Z(I)=Y(I)
C 23 CONTINUE
C P25=PCTILE(Z,KP,.25)
C P75=PCTILE(Z,KP,.75)
C P50=PCTILE(Z,KP,.50)
C YMIN=2.5#P25-1.5#P75
C YMAX=2.5#P75-1.5#P25
C IF(Z(1).GT.YMIN) YMIN=Z(1)
C IF(Z(KP).LT.YMAX) YMAX=Z(KP)
C RETURN
C END

```

## LIST OF REFERENCES

1. Heidelberger, P. and Lewis, P. A. W., "Regression-Adjusted Estimates for Regenerative Simulations, with Graphics," Communications of the ACM, v. 24, pp 260-273, April 1981.
2. Cox, D. R. and Lewis, P. A. W., The Statistical Analysis of Series of Events, Mehteun & Co. LTD, 1966.
3. Lewis, P. A. W. and Uribe, L., The New Naval Postgraduate School Random Number Package LLRANDOM, Naval Postgraduate School Rept. NPS55-81-005, Monterey, California, 1981.
4. Davis, Harold T., Tables of the Higher Mathematical Functions, 1st ed., v. 1, Principia Press, Inc., 1933.
5. Miller, R. G., Jr., "The Jackknife--a Review," Biometrika 61, April 1974.

# INITIAL DISTRIBUTION LIST

	No. Copies
1. Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2. Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3. Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
4. Captain David G. Linnebur, USMC Headquarters Marine Corps (MPI 20) Washington, D.C. 20380	1
5. Professor P. A. W. Lewis, Code 55LW Naval Postgraduate School Monterey, California 93940	5
6. Commandant of the Marine Corps, Code RDJS-40 Headquarters Marine Corps Washington, D.C. 20380	1
7. Dean of Research, Code 012 Naval Postgraduate School Monterey, California 93940	1
8. Professor P. A. Jacobs, Code 55JC Naval Postgraduate School Monterey, California 93940	1
9. Dean of Information and Policy Sciences, Code 01 Naval Postgraduate School Monterey, California 93940	1
10. Professor D. P. Gaver, Code 55GV Naval Postgraduate School Monterey, California 93940	1

- |     |  |   |
|-----|--|---|
| 11. | Lt. Col. J. F. Mullane, Code 0309<br>Naval Postgraduate School<br>Monterey, California 93940                       | 1 |
| 12. | Captain Nick A. Sottler<br>Marine Corps Recruiting Station<br>4727 Wilshire Blvd.<br>Los Angeles, California 90010 | 1 |
| 13. | Captain H. Keith Sullivan<br>Officer Candidate School<br>MCDEC<br>Quantico, Virginia 22134                         | 1 |